

**University of Exeter**  
**Graduate School of Education**

**Exploring EFL Assessors' Assessment Literacy in an English Preparatory  
Programme in Kuwait**

Submitted by Fatma Ismail Mohamed  
To the University of Exeter  
As a Thesis for the Degree of Doctor of Education  
In Teaching English to the Speakers of Other Languages (EdD TESOL)  
In December 2020

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: Fatma Ismail Mohamed

## **ABSTRACT**

This exploratory study is conducted to understand, re-conceptualise, and possibly develop Teachers'/Assessors' Assessment Literacy (TAL/AAL) in an English Preparatory Programme (EPP) at one university in Kuwait. It has been observed that the evaluation practices adopted in the described context depend mainly on tests and that the adopted Assessment Practices (APs) therefore do not fulfil one primary purpose of assessment, which is supporting learning. Exploring TAL could provide an understanding of their assessment awareness and if their APs are informed by solid assessment knowledge or not. Understanding the source of existing APs can guide assessment improvement and development in the described context.

In order to explore TAL, different data collection tools were used to collect qualitative data from eight assessors about their Assessment Literacy (AL). The research used an open-ended questionnaire, an open-ended checklist, a semi-structured interview, an open-ended report, and an unstructured interview to explore all components of AL suggested by Teacher Assessment Literacy in Practice (TALiP) framework (Xu & Brown, 2016). It was found that the participants' pre-and in-service training had not prepared them theoretically or practically for their assessment tasks. Their assessment knowledge base has not been developed through accredited channels or on the job. The participants were not confident with most of the components of the knowledge base proposed by the TALiP framework. However, they did show full awareness of assessment complexities in their teaching context. They were able to distinguish appropriate and inappropriate practices and their washback on learning. Their beliefs about assessments also resonated with recommended APs in the AL literature. It is recommended that the participants should receive theoretical and practical training in the assessment knowledge base. If they had a more substantial assessment knowledge base, they could have revealed better solutions to uncertainties surrounding them, and they could have been more precise in specifying their assessment needs.

## **Acknowledgement**

I would like to express my deepest gratitude to my first supervisor, Dr Philip Durrant, for his professional support, care, and patience with me throughout the whole research. He provided me with an excellent research experience. I had the honour of being Dr Philip's student and having him supervise my thesis. I would have never been able to finish my thesis without his guidance. I would also like to express my warm gratitude to my second supervisor, Dr Susan Riley, for her support and guidance throughout the programme. She was always there whenever I needed help. Special thanks to all my professors, whose professional attitude enriched my experience as a student in the EdD programme.

## **Dedication**

I dedicate my thesis work to my beloved husband, who supported me throughout my studies. I owe him every moment of success. If it were not for him, I would not have accomplished that much. I also dedicate my work to the soul of my beloved father. I wished he would have been by my side to witness the harvest of his support to me throughout my academic life. May his soul rest in peace. A special feeling of gratitude to my four daughters for being patient with me throughout the entire programme.



## List of Contents

<b>Abstract</b>	<b>2</b>
<b>Acknowledgement</b>	<b>3</b>
<b>Dedication</b>	<b>4</b>
<b>List of Contents</b>	<b>5</b>
<b>List of Tables</b>	<b>13</b>
<b>List of Figures</b>	<b>15</b>
<b>List of Abbreviations</b>	<b>16</b>
<b>Chapter One: Introduction</b>	<b>18</b>
1.1 Nature of the Problem	18
1.2 Description of the Research Context	20
1.2.1 General Description of the Context	20
1.2.2 The Learners	21
1.2.3 The Syllabus	21
1.2.4 The Teachers	21
1.2.5 The Adopted Assessment Approach	22
1.2.6 The Participants (The Assessors)	22
1.2.7 The Relation between the Researcher and the Participants	23
1.3 The Rationale of the Study	23
1.3.1 Contextual Rationale	23
1.3.2 Theoretical Rationale	25
1.3.3 Professional Rationale	26
1.4 Significance of the Study	28
1.5 Research Purpose and Questions	29

1.6	The Organisation of the Study	31
<b>Chapter Two: Literature Review</b>		<b>32</b>
Introduction		32
2.1	Definitions of Teachers Assessment Literacy (TAL) and Language Teachers Assessment Literacy (LTAL)	32
2.2	Research on Teachers' Assessment Literacy (TAL)	36
2.2.1	Research on Components of Assessment Literacy	37
2.2.2	Research on Measurements of Assessments Literacy	43
2.2.3	Research on Language Assessment Education	45
2.2.3.1	<i>Assessment Courses</i>	46
2.2.3.2	<i>Assessment Training Programs, Tools &amp; Recourses</i>	46
2.2.3.3	<i>Teachers' Assessment Training Needs</i>	47
2.2.4	Research on Understanding/Developing Assessment Literacy in Practice	49
2.3	Language Assessment Literacy Models/Frameworks	50
2.3.1	A Five Component Model (Brindley, 2001)	50
2.3.2	A Skill, Knowledge, and Principles Model (Davies, 2008)	51
2.3.3	A Practice, Principles, and Contexts Model (Fulcher, 2012)	51
2.3.4	LAL Stakeholders Profile Model (Taylor, 2013)	52
2.3.5	Teacher Assessment Literacy in Practice (TALiP) Model (Xu & Brown, 2016)	54
2.4	The Knowledge Base of Language Teacher Assessment Literacy	59
2.4.1	Disciplinary Knowledge & Pedagogical Content Knowledge	60
2.4.2	Knowledge of Assessment Purposes, Content, and Methods	62
2.4.2.1	<i>Theoretical Knowledge of Assessment Purposes, Content, and Methods</i>	62

2.4.2.2	<i>Practical Knowledge of Assessment Purpose, Content, and Methods</i>	68
2.4.3	Knowledge of Assessment Grading	73
2.4.4	Knowledge of Providing Feedback	76
2.4.5	Knowledge of Self-and Peer-Assessments	77
2.4.6	Knowledge of Assessment Interpretation and Communication	78
2.4.7	Knowledge of Assessment Ethics	79
2.5	Empirical Studies Conducted on Exploring Teachers' Assessment Literacy	81
Conclusion		83
<b>Chapter Three: Research Methodology</b>		<b>86</b>
Introduction		86
3.1	Research Paradigm	86
3.2	Research Methodology	88
3.3	Data Collection Methods	90
3.4	Data Collection Tools	93
3.4.1	A Questionnaire with Open-Ended Questions	94
3.4.2	A Structured Interview with an Open-Ended Checklist	95
3.4.3	A Semi-Structured Interview with Open-Ended Questions	101
3.4.4	A Structured Interview with an Open-Ended Report	102
3.4.5	An Unstructured Interview	103
3.5	Research Procedures	104
3.5.1	Sampling Procedures	104
3.5.2	Data Collection Procedures	107
3.6	Data Analysis	109

3.6.1	Analysis Method Conducted on Data Collected from the Open-Ended Questionnaire	110
3.6.2	Analysis Method Conducted on Data Collected from the Open-Ended Checklist	111
3.6.3	Analysis Method Conducted on Data Collected from the Semi-Structured Interview	113
3.6.4	Analysis Method Conducted on Data Collected from the Report	114
3.6.5	Analysis Method Conducted on Data Collected from the Un-Structured Interview	115
3.7	Research Trustworthiness	116
3.8	Ethical Considerations	118
<b>Chapter Four: Findings and Discussion</b>		<b>120</b>
Introduction		120
4.1	Findings Related to the Research Questions	120
4.1.1	Findings Related to Question 1. A	120
4.1.1.1	<i>Pre-Service Preparation</i>	120
4.1.1.2	<i>In-Service Preparation</i>	121
4.1.1.3	<i>Teachers' Effort</i>	122
4.1.2	Findings Related to Question 1. B	122
4.1.2.1	<i>Assessment Role</i>	123
4.1.2.2	<i>Selection Criteria</i>	123
4.1.3	Findings Related to Question 2. A	123
4.1.3.1	<i>Disciplinary &amp; Pedagogical Content Knowledge</i>	124
4.1.3.2	<i>Knowledge of Assessment Purposes, Content, and Methods</i>	125
4.1.3.3	<i>Knowledge of Assessment Grading</i>	129
4.1.3.4	<i>Knowledge of Assessment Feedback</i>	131
4.1.3.5	<i>Knowledge of Peers- and Self-Assessments</i>	133

4.1.3.6	<i>Knowledge of Assessment Interpretation and Communication</i>	135
4.1.3.7	<i>Knowledge of Assessment Ethics</i>	137
4.1.4	Findings Related to Question 2. B	141
4.1.4.1	<i>Need for Disciplinary Knowledge &amp; Pedagogical Content Knowledge</i>	141
4.1.4.2	<i>Need for Knowledge of Assessment Purposes, Content, and Methods</i>	142
4.1.4.3	<i>Need for Knowledge of Assessment Grading</i>	144
4.1.4.4	<i>Need for Knowledge of Providing Formative Feedback</i>	145
4.1.4.5	<i>Need for Knowledge of Peer- and Self-Assessments</i>	146
4.1.4.6	<i>Need for Knowledge of Assessment Interpretation and Communication</i>	147
4.1.4.7	<i>Need for Knowledge of Assessment Ethics</i>	148
4.1.5	Findings Related to Question 3. A	149
4.1.5.1	<i>Assessment Purpose</i>	149
4.1.5.2	<i>Assessment Sources</i>	149
4.1.5.3	<i>Assessment Methods</i>	150
4.1.5.4	<i>Assessment Decision-Making</i>	150
4.1.5.5	<i>Assessment Design Process</i>	150
4.1.5.6	<i>Assessment of Language Skills</i>	150
4.1.5.7	<i>Assessment Administration</i>	151
4.1.5.8	<i>Assessment Grading</i>	151
4.1.5.9	<i>Peer- and Self-Assessments</i>	152
4.1.5.10	<i>Assessment Feedback</i>	152
4.1.5.11	<i>Assessment Interpretation and Communication</i>	152
4.1.5.12	<i>Assessment Reliability and Validity</i>	152
4.1.5.13	<i>Assessment Fairness and Ethical Procedures</i>	153
4.1.6	Findings Related to Question 3.B	153

4.1.6.1	<i>Beliefs about Assessment Purpose</i>	153
4.1.6.2	<i>Beliefs about Assessment Sources</i>	153
4.1.6.3	<i>Beliefs about Assessment Methods</i>	154
4.1.6.4	<i>Beliefs about Assessment Decision-Making</i>	154
4.1.6.5	<i>Beliefs about the Assessment Design Process</i>	155
4.1.6.6	<i>Beliefs about Assessment of Language Skills</i>	155
4.1.6.7	<i>Beliefs about Assessment Administration</i>	156
4.1.6.8	<i>Beliefs about Assessment Grading</i>	157
4.1.6.9	<i>Beliefs about Peer- and Self-Assessments</i>	157
4.1.6.10	<i>Beliefs about Assessment Feedback</i>	157
4.1.6.11	<i>Beliefs about Assessment Interpretation and Communication</i>	158
4.1.6.12	<i>Beliefs about Assessment Reliability and Validity</i>	158
4.1.6.13	<i>Beliefs about Assessments Fairness and Ethical Procedures</i>	158
4.1.7	Findings Related to Question 4	158
4.1.7.1	<i>Assessment Purpose</i>	159
4.1.7.2	<i>Assessment Decision-Making</i>	159
4.1.7.3	<i>Assessment Source</i>	159
4.1.7.4	<i>Assessment Methods</i>	160
4.1.7.5	<i>Assessment Design Process</i>	160
4.1.7.6	<i>Assessment of Language Skills</i>	160
4.1.7.7	<i>Assessment Administration</i>	161
4.1.7.8	<i>Assessment Grading</i>	162
4.1.7.9	<i>Peer- and Self- Assessments</i>	162
4.1.7.10	<i>Assessment Feedback &amp; Result Communication</i>	162
4.1.7.11	<i>Assessment Reliability &amp; Validity</i>	163
4.1.7.12	<i>Assessment Ethics and Fairness</i>	163
4.1.8	Findings Related to Question 5	163
4.1.8.1	<i>Experience with the Open-Ended Checklist</i>	164

4.1.8.2	<i>Experience with the Semi-Structured Interview</i>	165
4.1.8.3	<i>Experience with the Open-Ended Report</i>	165
4.1.8.4	<i>Experience with the Whole Research</i>	166
4.1.8.5	<i>Research Impact on Future Development in Assessment Literacy</i>	166
4.2	Discussion of the Research Findings	167
4.2.1	Discussion of Question 1.A Findings	167
4.2.2	Discussion of Question 1.B Findings	168
4.2.3	Discussion of Question 2.A Findings	169
4.2.4	Discussion of Question 2.B Findings	172
4.2.5	Discussion of Question 3.A Findings	174
4.2.6	Discussion of Question 3.B Findings	175
4.2.7	Discussion of Question 4 Findings	178
4.2.8	Discussion of Question 5 Findings	179
Conclusion		181
<b>Chapter Five: Conclusion</b>		<b>182</b>
5.1	Summary of the Findings	182
5.2	Research Implications	185
5.3	Recommendations	186
5.4	Research Contribution to Knowledge	187
5.5	Merits of the Study	189
5.6	Limitations of the Study	190
<b>Bibliography</b>		<b>192</b>
<b>Appendices</b>		<b>212</b>
Appendix 1: First Research Tool - Open-Ended Questionnaire		212

Appendix 2: Second Research Tool - Open-Ended Checklist	214
Appendix 3: Third Research Tool - Semi-Structured Interview	232
Appendix 4: Fourth Research Tool - Open-Ended Report	233
Appendix 5: Fifth Research Tool - Unstructured Interview	234
Appendix 6: Ethics Application Form	235
Appendix 7: Certificate of Ethical Approval	242
Appendix 8: Participant Information Sheet	243
Appendix 9: Consent Form	247
Appendix 10: Mirjana's Transcriptions	249
Appendix 10.1: Transcription of Mirjana's Open Ended- Questionnaire	249
Appendix 10.2: Transcription of Mirjana's First Structured Interview (Checklist)	254
Appendix 10.3: Transcription of Mirjana's Semi-Structured Interview	287
Appendix 10.4: Transcription of Mirjana's Second Structured Interview (Report)	302
Appendix 10.5: Transcription of Mirjana's Completed Report	313
Appendix 10.6: Transcription of Mirjana's Unstructured Interview	325
Appendix 11: Analysis of Data Collected from The Open-Ended Questionnaire	329
Appendix 12: Analysis of Data Collected about the Participants' Confidence Level with Assessment Knowledge Base	336
Appendix 13: Analysis of Data Collected about the Participants' Need for the Proposed Assessment Knowledge Base	355
Appendix 14: Analysis of Data Collected about the Participants' Context-Adopted Assessment Practices	380
Appendix 15: Analysis of Data Collected about the Participants' Beliefs of their Context-Adopted Assessment Practices	389
Appendix 16: Analysis of Data Collected about the Participants' Suggestions for Better Assessment Practices	404
Appendix 17: Analysis of Data Collected about the Participants' Experience with the Research Activities	426



## **List of Tables**

<b>Table 2.1</b>	Governments and Research-Based Assessment Organizations which Specified Assessment Competencies
<b>Table 2.2</b>	Professional Attempts Adopted by Teacher Accreditation and Certification Organizations to Set Standards for TAL
<b>Table 2.3</b>	Eight LTAL Measurement Instruments
<b>Table 2.4</b>	Types of Assessments
<b>Table 3.1</b>	The Participants' Pseudonyms, Gender, Qualification and Experience
<b>Table 4.1</b>	The Participants' Responses to and Comments on Five Components Associated with Disciplinary and Pedagogical Knowledge
<b>Table 4.2</b>	The Participants' Responses to and Comments on Eight Components Associated with Theoretical Knowledge of Assessment Purposes, Content, and Methods
<b>Table 4.3</b>	The Participants' Responses to and Comments on Sixteen Skills Associated with Knowledge of Assessment Purposes, Content, and Methods
<b>Table 4.4</b>	The Participants' Responses to and Comments on Six Components Associated with Theoretical Knowledge of Assessment Grading
<b>Table 4.5</b>	The Participants' Responses to and Comments on Six Skills Associated with Knowledge of Assessment Grading
<b>Table 4.6</b>	The Participants' Responses to and Comments on Three Components Associated with Theoretical Knowledge of Assessment Feedback
<b>Table 4.7</b>	The Participants' Responses to and Comments on Eleven Skills Associated with Knowledge of Assessment Feedback
<b>Table 4.8</b>	The Participants' Responses to and Comments on Two Components Associated with Theoretical Knowledge of Peers- and Self-Assessments
<b>Table 4.9</b>	The Participants' Responses to and Comments on Five Skills Associated with Knowledge of Peers- and Self-Assessments

<b>Table 4.10</b>	The Participants' Responses to and Comments on Two Components Associated with Theoretical Knowledge of Assessment Interpretation and Communication
<b>Table 4.11</b>	The Participants' Responses to and Comments on Ten Skills Associated with Knowledge of Assessment Interpretation and Communication
<b>Table 4.12</b>	The Participants' Responses to and Comments on Four Components Associated with Theoretical Knowledge of Assessment Ethics
<b>Table 4.13</b>	The Participants' Responses to and Comments on the Sixteen Skills Associated with the Knowledge of Assessment Ethics
<b>Table 4.14</b>	Components of Disciplinary and Pedagogical Knowledge
<b>Table 4.15</b>	Theoretical Components of Assessment Purpose, Content and Methods
<b>Table 4.16</b>	Skills Associated with Assessment Purpose, Content, and Methods
<b>Table 4.17</b>	Components of Assessment Grading Knowledge
<b>Table 4.18</b>	Components of Assessment Feedback Knowledge
<b>Table 4.19</b>	Components of Peer- and Self-Assessments Knowledge
<b>Table 4.20</b>	Component of Assessment Interpretation and Communication Knowledge
<b>Table 4.21</b>	Components of Assessment Ethics Knowledge

## **List of Figures**

<b>Figure 2.1</b>	Language Assessment Literacy Model (Fulcher, 2012)
<b>Figure 2.2</b>	LAL Profile for Classroom Teachers (Taylor, 2013)
<b>Figure 2.3</b>	Conceptual Framework of TALIP Model (Xu & Brown, 2016)
<b>Figure 3.1</b>	Assessment Knowledge Base as Presented in the Checklist.
<b>Figure 3.2</b>	Thematic Categorization and Coding of Data Collected from the Questionnaire.
<b>Figure 3.3</b>	Thematic Categorization and Coding of Data Collected from the Checklist.
<b>Figure 3.4</b>	Thematic Categorization and Coding of Data Collected from the Semi-Structured Interview
<b>Figure 3.5</b>	Thematic Categorization and Coding of Data Collected from the Unstructured Interview

## **List of Abbreviations**

<b>Abbreviations</b>	<b>Meaning</b>
AAL	Assessors' Assessment Literacy
ACACA	The Australian Curriculum & Assessments & Certificate Authorities
AE	Academic English
AEA	Association for Educational Assessment
AERA	The American Educational Research Association
AFT	American Federation of Teachers
AFL	Assessment for Learning
AL	Assessment Literacy
ALT	Assessment Literacy Inventory
APs	Assessment Practices
APA	American Psychological Association
API	Assessment Practices Inventory
ARG	Assessment Reform Group
CALI	Classroom Assessment Literacy Inventory
CEFRL	Common European Framework of Reference for Languages
DA	Dynamic Assessment
EAP	English For Academic Purposes
EFL	English as a Foreign Language
ELA	English Language Assessment
EPP	English Preparatory Programme
INTASC	Interstate Teachers Assessment and Support Consortium

JAC	Joint Advisory Committee
JCSEE	Joint Committee for Standards on Educational Evaluation
JEAP	Journals of English for Academic Purposes
LA	Language Assessment
LAL	Language Assessment Literacy
LTAL	Language Teachers Assessment Literacy
NBPTS	National Board for Professional Teaching Standards
NCATE	National Council for the Accreditation of Teacher Education
NCME	National Council on Measurement in Education
NEA	National Education Association
RALT	Revised Assessment Literacy Inventory
TAL	Teacher Assessment Literacy
TALiP	Teacher Assessment Literacy in Practice
TALQ	Teachers Assessment Literacy Questionnaire
UK	United Kingdom
USA	United States of America

# **Chapter One: Introduction**

## **1.1 Nature of the Problem**

English Language Assessment (ELA) has become a crucial element in our modern life. It became endemic in employment, international mobility, and, most importantly, in educational institutions (Fulcher, 2010). It is nearly an essential component of all educational levels, including elementary, secondary, and tertiary education. The purpose of ELA is to help assessors make inferences about candidates' language abilities and make decisions based on these inferences (Weigle, 2002). However, language assessment Impact on learning can be harmful if the assessment content and techniques are at variance with the course objectives, mostly nowadays assessments are not seen as measuring tools only; they are expected to foster educational growth and enable candidates to benefit from the assessment experience (Davidson & Lynch, 2002).

In tertiary education, the widespread expansion of English medium instruction in universities worldwide has led to more focus on teaching English for Academic Purposes (EAP) (Wachter & Maiworm, 2014). Universities use English Preparatory Programmes (EPPs) for undergraduate learners attending universities after high school to teach EAP to assist learners' study and research in that language (Hyland, 2006). These programmes provide learners with skills and strategies that enable them to function successfully in academic contexts and perform specific academic tasks, such as writing academic papers, reading different genres, delivering academic presentations, and participating in meetings and conferences using English (Hyland, 2002). In these programmes, assessments play a significant role, as they act as filters to acceptance to universities and colleges or pass cards to continuation in such academic environments. Decisions taken in EPPs are high-stakes for would-be students (Flowerdew & Peacock, 2001).

Assessments at EPPs are also meant to fill gaps and address criticism directed at large-scale international tests, which are also used as evidence for learners' language proficiency in academic contexts (Schmitt & Hamp-Lyons, 2015). Time constraints and the need to assess candidates from a wide range of educational backgrounds, in addition to their reductionist nature, make large-

scale international tests present a limited sample of EAP genres. On the other hand, EPPs can better represent the construct of academic language in use because they are context-bound and related to particular course objectives and goals meant to develop EAP (Schmitt & Hamp-Lyons, 2015). They are meant to take account of students' experiences, needs, expectations, and preferences (Bruce, 2011). Assessments taking place in EPPs "are more likely to allow making valid assumptions about test-takers" (O'Sullivan, 2011:270). It is claimed that EPPs have the potential for innovation in language assessment design and improved understanding of the construct of EAP (Schmitt & Hamp-Lyons, 2015). Assessments in EAP programmes are expected to be more directed to achievement assessment instead of proficiency testing. Classroom-based and course-work style achievement assessments provide assessors with opportunities to gather evidence on what learners can and cannot do with course objectives.

However, in-house assessments that are designed and administrated in EPPs are sometimes constrained by political, social, or financial issues (Schmitt & Hamp-Lyons, 2015). Many EFL teachers/assessors in EPPs use large-scale proficiency style exams to assess learners at the end of their programmes to meet requirements imposed on them even though teaching and ongoing assessments are directly related to material or textbooks that are meant to develop learners' competence in accomplishing academic tasks. Assessors need to balance their beliefs, learners' rights, and programmes' needs to propose fair, valid assessments. To perform such a challenging task, EFL teachers/assessors need to be armed with solid Assessment Literacy (AL). Teacher Assessment Literacy (TAL) should not only involve a set of knowledge and skills available in testing reference books and training programs that assessors might apply blindly. TAL should be derived from experience, academic knowledge, skills, beliefs, and critical reflection on daily assessment practices. All these components of TAL would interact to form their identities as assessors and enable them to make compromises that satisfy their students, context policies, and most important their beliefs about assessments (Willis et al., 2013; Xu & Brown, 2016).

Such a form of AL can only be explored in its real contexts to be revealed to EAP communities and the wider assessment community to define EAP

assessment constructs and appropriate assessment practices bound to EAP contexts, since "Efforts to develop assessment literacy within EAP community itself are often fragmented, stop/start affairs" (Manning, 2013). Manning (2014) argued for a clear understanding of Assessors Assessment Literacy (AAL). Therefore, I intend to conduct an exploratory interpretive study in an EPP in one private university in Kuwait with the aim of understanding, reconceptualising and possibly developing TAL/AAL. In the following sections, I will present a brief description of the research context, in addition to the rationale, significance, and purpose of the study. The chapter will end with the proposed research questions, which the research will try to answer.

## **1.2 Description of the Research Context**

This section provides a brief description of the current research context; however, it does not reveal details about the context for two reasons. For ethical reasons, the context and the participants' identity need to be protected. Kuwait is a small country with a limited number of universities and English foundation programmes, and teachers working in this field know each other. Giving a detailed description of the context can reveal their identity and the programme identity.

### **1.2.1 General Description of the Context**

The context of the research study is an EPP in one private university in Kuwait. Universities in Kuwait require learners to provide evidence of English language proficiency before starting their undergraduate courses because all courses are taught in English. If learners fail to fulfil this requirement, they join a preparatory year, in which they study English in an EPP established at the university. The programme lasts for a whole academic year; it is divided into two levels and given in two semesters of fifteen weeks each. Success in level one allows learners to join level two. Once learners complete the programme successfully, they can start their undergraduate courses, which they have previously selected. The EPP aims to help the learners develop linguistically and academically to perform competently when joining their academic disciplines. Learners need to show competence in English and awareness of academic



features before joining their majors through different assessments conducted throughout the year.

### **1.2.2 The Learners**

All learners joining an EPP are Kuwaiti high school graduates aged between 18 and 19. Most of them received their education at public government schools, in which English is taught as a foreign language in a traditional way that emphasises grammar instruction and practises, in addition to vocabulary memorisation. Little attention is paid to language use and communication, which is why learners find the programme challenging, especially at the beginning, as the programme employs a learner-centred approach that underlines inductive, cooperative, and task-based learning. Both males and females learners join the programme; however, they receive their English language courses in segregated classes.

### **1.2.3 The Syllabus**

Learners studying at an EPP take three language modules throughout the entire academic year. They are introduced to academic reading, listening, and presentations skills, in addition to academic vocabulary, in one module. In the second module, the learners are introduced to academic writing; they practise writing sentences, academic paragraphs, and essays. They practise using different transitions for cohesion, in addition to summarising, paraphrasing, and referencing. In the third module, the learners focus on specific points of English grammar to increase their understanding of grammar rules to support their reading, listening, and writing performance. They are also exposed to different grammatical error codes, which they are required to use and interpret while they practise editing and drafting in the writing module. The programme adopts varied progressive and integrated teaching methods, approaches, and techniques.

### **1.2.4 The Teachers**

Teachers working in these contexts have varied backgrounds. They are not local teachers; in other words, they are not from Kuwait. Some are native-speakers of English, while others are not, but they are all experienced in teaching English as a Foreign Language (EFL). They worked in different countries and contexts before coming to Kuwait. They are all qualified academically; most of them hold master's degrees and have at least five years of teaching experience

before joining the programme. Each teacher is assigned one module to teach. It is a collaborative teaching process, where learners in each section receive language instructions from different sources and through different experiences. Teachers working in such programme are typically engaged in continuous professional development opportunities through annual workshops and seminars provided by the team leaders and supervisors. These workshops aim to emphasise the programme's teaching philosophy, enhance teaching practices, and engage teachers in collaborative professional development practice. The teachers are normally excluded from assessment preparation, design, and construction. The assessment tasks are assigned to the modules' team leaders, who are excluded from teaching and daily in-class contacts with the learners.

### **1.2.5 The Adopted Assessment Approach**

In each semester, learners sit for mid-term and final exams. The exams in EPPs focus on assessing vocabulary, grammar, reading comprehension skills, and writing skills. Items assessing grammar and vocabulary are in the form of multiple-choice questions. Reading comprehension is assessed through a given reading text followed by multiple-choice or true/false questions. Writing is assessed through questions assessing sub-writing skills. In addition, learners are asked to write paragraphs or short essays, on which they are trained in advance. Speaking skills are assessed through oral presentations, which learners are expected to present in class. The topic of each presentation is given in advance to the learners to prepare and display in class. Learners' listening skills are assessed through listening tests that follow audio-recorded texts, to which learners are required to listen and answer multiple-choice, true/false, and fill in the blank questions. Learners' attendance and participation in activities are taken into consideration while calculating the overall passing grade for each semester. Upon passing the first semester, learners join semester two, in which they are exposed to more advanced materials, but the same assessment scheme is used.

### **1.2.6 The Participants (The Assessors)**

The participants in the current research study were the modules' team leaders, who were selected by the programme lead to perform all required assessment tasks. They were engaged in assessment design and administration guided by the programme's adopted assessment philosophy, strategies, goals, and practices. All participants/assessors were previously engaged in teaching

before being selected as assessors. The assessors' group was rather a homogeneous one; they shared many characteristics. The assessors were all involved in teaching in the described context for more than five years; they were academically prepared to teach EFL/ESL. As they all work at the EPP, they were given the same assessment tasks and followed the same assessment guideline provided by the programme director. They received the same institutional periodic training focusing on the institutional adopted teaching and learning philosophies.

### **1.2.7 The Relation between the Researcher and the Participants**

I worked with the assessors in the same university; however, we did not work in the same programme. They worked in the EPP; I worked in an administrative department performing an administrative role focusing on reviewing assessment guidelines related to invigilation, storage, and filling for the whole university. I used to work with them as an EFL instructor five years ago. I did not have any authority over them; their participation in the current research was voluntarily guided by their interest in the topic, which I presented for them. Our relation was framed by collegiality, mutual respect, and common interest in the same topic. We aimed to develop our understanding of the observed problem and possibly develop our assessment literacy at least on the conceptual level; this opportunity helped us reflect on our current assessment literacy to take further development steps on the practical level.

## **1.3 The Rationale of the Study**

### **1.3.1 Contextual Rationale**

Reviewing the assessment scheme adopted in the described EPP, it is found that assessments depend mainly on tests. Learners sit for mid-term and final exams that form 50% of the total assessment grade. The exams focus on assessing different language skills using close-ended items with very little chance for learners producing a real-life output similar to the ones they are expected to produce in their academic disciplines. Even with the productive skills, such as writing and speaking, the learners' presented tasks, on which they would be assessed, are prepared in advance through guided instructions, and the output of the tasks is either controlled, practised, or memorised.

Working in three different EPPs in three different Kuwait universities, I have observed that nearly the same assessment scheme is adopted in these contexts. The adopted assessment practices do not fulfil one of the primary purposes of assessment, supporting learning, as emphasised by Djoub (2017). The assessment items do not engage learners in problem-solving, predicting, reflecting, creating, interpreting, investigating, or applying knowledge to new situations. They are limited to engaging learners into simple learning experiences, such as recalling, answering, recording, organising, making choices, and transforming (Cameron, Tate, Macnaughton, & Politano, 1998). The adopted assessments do not include authentic performance tasks that require learners to use language to perform academic tasks, such as note-taking, interpreting and interacting with academic texts, analysing spoken or written texts, writing argumentative assignments, or writing reports. Learners are not given a chance to comment on their work or their peers' work. Moreover, no feedback is given to the learners to close the gap between their performance after assessment and the target performance (Sadler, 1989). All that the learners receive are scores, which can foster the fear of failure, depress creativity, weaken students' interest, and emphasise quantitative aspects of learning as emphasised by Bulter & Nisan (1986).

Moreover, other empirical studies conducted in similar programmes in different Middle East countries reported similar assessment schemes like the one adopted in the described context. For example, in a study conducted by Elmahjoub (2017) in the Libyan context, it was found that traditional modes of testing dominated in assessing language in different language programmes. He emphasised that tests focused on measurable outcomes and neglected unmeasurable objectives, which led to narrowing taught material and made teachers teach to the tests. In another study conducted by Efeotor (2017) in the Saudi context, he concluded that the assessment practices adopted in the studied foundation programme failed to carry out any formative function; there were no steps taken to make learners learn from the assessment experience whether during or after assessment. Students receive feedback in the form of scores or grades. He also found that it was common for students to memorise items and recall them in the test. Exploring research on LA in the Middle East and

North Africa, it is believed that the assessment procedures tend to be more into the summative side (Coombe et al., 2017). Coombe believes that test scores are still the leading indicators and, in many cases, the only one for assessing learning outcomes. Shohamy et al. (2017) emphasised that tests in the Middle East are used to measure specific pre-determined linguistic knowledge. These research findings support the observed problem in the studied context.

Based on the previous observation, it can be concluded that the assessment practises (APs) in the described context do not reflect the kind of assessments that are meant for developing EAP and consequently assessing learners' ability in this area. It is assumed that EAP constructs, which need to be addressed in assessments, are broader than just language ones (Chiu, 2015; Lea & Street, 2000). Therefore, I believe that the kind of inferences proposed by the described programme lack face, construct, and content validity because of "low item-objective congruence" (Hambleton, 1994: 23) and variance between the programme objectives and assessment content. Accordingly, the decisions made based on the assessment inferences cannot predict how well learners will function in their academic disciplines. This assessment approach makes the APs adopted in the described context subject to ethical questions. According to Shohamy (2001), test/assessment ethicality is about their role in societies and their misuse. Assessment ethicality is a call for quality assessments that require meeting various criteria of validity.

### **1.3.2 Theoretical Rationale**

Exploring Journals of English for Academic Purposes (JEAP), it is mentioned that few EAP practitioners are involved in debates about what constitutes EAP assessments. "The development of EAP assessment theory has been left in the hand of language testers even though it is EAP practitioners who devise and administer most of the EAP assessment going on around the world" (Schmitt, Hamp-Lyons, 2015:3). Practices, perceptions, and beliefs of EAP practitioners about assessment are under-explored, under-defined, and under-theorised; therefore, it can be claimed that EAP practitioners have contributed little to theory-building or establishment of an understanding of EAP assessment practices. Most of the research conducted on EAP assessments was associated with standardised exams, while research conducted on in-house EAP

assessments, such as the ones conducted in different EPPs, is under-represented in an international publication like JEAP; such studies, if any, are heard only in local publications and websites read by fellow practitioners rather than language testers. It can be concluded that "voices of teachers as EAP assessment developers are not being heard outside their community" (Schmitt, Hamp-Lyons, 2015:3). This situation has resulted in two problems. One is isolation between EAP communities and language testing and assessment communities, and the other is insufficient Knowledge of EAP practitioners' AL. Research conducted on EAP assessment practices can contribute to understanding EAP assessment practices and the development of EAP constructs.

Therefore, I would argue that a gap in the literature of EAP practitioners' AL needs to be addressed through exploring and understanding their practices, perceptions, and beliefs of in-house EAP assessment, a type of assessment that is context-bound and could be understood only in its contexts through the eyes of those involved in it. Taking into consideration that mainstream assessment textbooks focus mainly on large-scale tests and general classroom assessments; I have come across very few resources that offer guidance to EAP practitioners designing EAP in-house assessments.

### **1.3.3 Professional Rationale**

Exploring the field of Teacher Assessment Literacy (TAL), it was found that the field was explored from different perspectives. Earlier studies focused on the assessment knowledge base to define TAL (Brookhart, 1999, 2011; Popham, 2009, 2011; Stiggins, 1991, 1999). Another stream of studies focused on measurement tools of TAL to judge them as being literate or illiterate (Campbell et al., 2002; Jarr, 2012; Mertler, 2003; Plake et al., 1993). The third group of studies focused on teachers' assessment education (Bailey & Brown, 1995; Brown & Bailey, 2008; Greenberg & Walsh, 2012), teachers' continuous assessment development (Graham, 2005; Sato, Wei & Darling Hammond, 2008), and teachers' assessment needs to provide them with appropriate training based on their needs (Fulcher, 2012; Hasselgreen, Carlsen, & Helness, 2004; Vogt & Tsagari, 2014). Some of these studies viewed Language Assessment Literacy

(LAL) as a kit that could be taken by teachers to be used in any situation. They ignored the behavioural and contextual aspects that can interfere in shaping TAL.

Although there were different research attempts that adopted a socio-cultural view of TAL, I believe more attention should be paid to the socio-cultural aspect of TAL and the dynamic and complicated nature of TAL. TAL includes teachers' perceptions, emotions, and beliefs about what they were exposed to through formal instructions during their formal preparation as teachers and experience as students. Teachers carry all these inputs about assessment with them when they work in their teaching contexts, where they are faced with assessment policies, philosophies, culture, guidelines that are meant to guide their APs. At this point, teachers can face three scenarios; they can abide by what is imposed on them blindly without opposing and therefore, their contexts assessment philosophy will inform their APs; this will result in discrepancies between teachers' perceptions and beliefs of assessment and their contexts', in addition to fossilising their AL and professional development. The second scenario would be teachers' opposition to their context-adopted philosophy, resulting in dissatisfaction or teachers' burnout. The last scenario would be teachers' tendency to take their literacy to an upper level, where they self-develop themselves through reflective practice to reach a compromise between their perceptions and beliefs about assessments and their macro and micro-teaching contexts to be able to form their assessment identity. This identity is in a continuous evolution state due to changes in their knowledge, beliefs, contexts (Xu & Brown, 2016). This approach will result in continuous development in TAL, which will enable them to provide learners with ethical, fair assessments.

This socio-cultural view of AL gives teachers a voice and involvement in APs (Shohamy, 2001), it also bridges the gap between theory and practice; besides, it shows that AL is dynamic and changeable from one context to another. To implement such a scenario or model, teachers involved in APs need to participate in research that makes them utter their own understanding/beliefs and reflect on their practices. This participation enhances teachers' understanding of their practices, gives them a voice and role in further assessment transformation, and provides literature with a holistic view of how AL is constructed.

## **1.4 Significance of the Study**

Based on the contextual observation and explored theoretical and empirical studies conducted on TAL, I believe that conducting an exploratory interpretive study in the described context to explore, understand, reconceptualise, and possibly develop TAL is essential for different reasons.

First, on the practical level, such a study would address and explain the observed problem (Denicolo & Becker, 2012), as it could provide an understanding and awareness of TAL. This awareness would show whether the assessors' APs are informed by their assessment philosophies or their context-adopted assessment philosophies. Understanding of the source of limited APs adopted in the studied programme can guide transformative actions in further studies to enhance APs in this programme; it will clarify whether actions need to be taken on the teachers' professional development level or administrative level. Moreover, these studies can give teachers/assessors a voice and contribution in this intended transformation.

Second, on the academic level, conducting the study can add to EAP assessment knowledge conceptually and empirically and address a gap in the literature (Denicolo & Becker, 2012). The study's output can act as an attempt to bridge the gap between EAP assessors and general language testers and assessors by publishing the findings in well-known journals and presenting them in recognised assessment conferences; this would expose the findings of the research for debate.

Third, I believe the study is worth doing for practical professional development reasons (Campbell & Collins, 2007; Gilbert, 2006). Having language teachers/assessors participate in such research will enable them to share their classroom assessment practices with others and develop a community of assessment learning (Nieto, 2010). It will also help them to assess their assessment knowledge, reflect on their assessment training and practices, utter their beliefs about them, evaluate their context-bound APs, and propose solutions to uncertainties and faced problems. All these research practices would contribute to the participants' professional development about assessment or



guide further professional development opportunities. It is crucial for stakeholders involved in APs to know why they assess, what they assess, and how to assess, in addition to the impact of their APs on individuals and societies to avoid negative washback of their APs, for which they are held ethically and professionally accountable. "Testing and assessments, in general, are more than a technical activity; they are also ethical enterprise" (Fulcher & Davidson, 2007: 1).

Finally, exploring TAL in the described program can act as a reference for further studies in similar programs, which could enable researchers to compare assessment literacies in different contexts of the same domain (EAP). This comparison could, in turn, provide the opportunity to investigate the claimed assumption by Xu & Brown (2016) regarding having different literacies in different contexts and the role of context in teachers'/assessors' assessment perceptions. The comparison can also allow researchers to develop similarities and differences between TAL in the different studied contexts to come up with general guidelines about EAP assessment practices and constructs. This guideline can be a reference for EAP assessors regarding their assessment tasks considering that I have not come across studies conducted in Kuwait, focusing on TAL with this perceptive.

## **1.5 Research Purpose and Questions**

The purpose of the current study is to understand, reconceptualise, and possibly develop eight EFL teachers'/assessors' assessment literacy in practice. This empirical research uses a conceptual framework of teacher assessment literacy in practice (TALiP) proposed by Xu & Brown (2016) as an operational model for researching and exploring TAL in an EPP in one private university in Kuwait. The study seeks answers to the following questions to fulfil the research purpose:

- 1.A How were the assessors prepared for their assessment tasks?
- 1.B How are the assessors involved in assessment tasks in their teaching context?
- 2.A How confident are the assessors with the assessment knowledge base suggested by the TALiP framework?
- 2.B What are the assessors' views regarding their need for different assessment knowledge base components suggested by the TALiP framework?

- 3.A How are assessments practised in the assessors' teaching context?
- 3.B What are the assessors' views about their context-adopted assessment practices?
4. How can the assessors negotiate between their views about assessments and their context-adopted assessment philosophies and practices?
5. How far have the assessors' assessment literacies been developed as a result of taking part in the research?

The research questions are meant to fulfil the following research objectives:

- a. Explore and understand the assessors' formal/informal assessment preparation during pre-service and in-service phases, in addition to their current involvement in assessment practices in their teaching context. This information will show how far the assessors were prepared for their assessment roles and duties, which they will share.
- b. Expose the assessors to the knowledge base suggested by the TALiP model, which they might be familiar with or not. If they are not, such exposure will add to their knowledge and open up topics for them to consider in their future APs and development. Most importantly, this would uncover the assessors' confidence with the proposed assessment knowledge base, which is considered an essential fundamental component of TAL (Xu & Brown, 2016).
- c. Elicit the assessors' beliefs about what is needed from the presented AL knowledge base as per their conceptions. This step is essential because not all assessment knowledge base would be required from different stakeholders with the same level (Taylor, 2013). Moreover, teachers' beliefs about what is needed or not needed as per their conceptions and contextual demands are essential components of the adopted TALiP model and one main component of TAL that is neglected (Deluca & Lam, 2014).
- d. Elicit information about the assessors' context-adopted APs. Knowledge of the macro and micro assessment constructs is another TAL level that needs to be uncovered because assessors' APs are affected by these domains (Crossman, 2007).

- e. Elicit the assessors' views about their context-adopted APs; this is an essential step because it would reveal their personal views, an area which has long been neglected in AL research (Hill et al. 2010) and another level of their TAL that needs to be explored.
- f. Allow the assessors to reflect on their context-bound assessment practices and compare them to their own beliefs, a standpoint, which they need to experience and an advanced level of their AL that contributes in forming their assessment identity (Woolfolk, Hoy, & Davies, 2008)
- g. Assist the assessors in making assessment-related decisions by engaging them in a collaborative professional dialogue in the study to develop solutions to conflicts that arose as a result of discrepancies between their views and their contextual assessment practices (Willis et al., 2013; Wyatt-Smith et al., 2010).
- h. Develop the assessors' assessment identity/literacy by helping them make use of the research experience through uttering areas of development they experienced as a result of taking part in the research and future areas of assessment development they would like to explore as a result of uncovering areas of weakness (Gottheiner & Siegel, 2012).

## **1.6 The Organisation of the Study**

The thesis is divided into five main chapters, including the introduction chapter. Chapter two will present the explored literature related to TAL. It will be followed by chapter three, in which the research methodology will be thoroughly presented. Chapter four will present the findings of the research. Finally, chapter five will sum up the whole research study by presenting the findings summary, recommendations, and study limitations.

## **Chapter Two: Literature Review**

### **Introduction**

This chapter presents the literature review conducted on TAL. The chapter aims to form a broad understanding of TAL and how it could be explored and developed. The outcome of the explored literature helped in choosing the research framework, shaping the research questions, and designing the research methodology. The chapter is divided into five main sections. Section one focuses on defining TAL and Language Teachers Assessment Literacy (LTAL). The second section presents research conducted in the AL field and how it passed through different stages; while the third section focuses on LTAL models designed to guide LTAL professional development. Section three ends with choosing one of the presented models/ frameworks as the theoretical framework, which will guide exploring the participants' AL. Section four focuses on presenting the components and sub-components of the AL knowledge base as specified by the adopted framework. Exploring LTAL knowledge base is essential, as it would guide the content of the adopted data collection tools designed to explore the participants' AL. Finally, section five presents a summary of similar empirical research studies conducted on LTAL.

### **2.1 Definitions of Teachers Assessment Literacy (TAL) and Language Teachers Assessment Literacy (LTAL)**

The term AL was coined by Stiggins (1991) based on the assessment competence standards developed by the American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), and the National Education Association (NEA) in 1990. In this early stage, AL referred to the knowledge and skills that stakeholders involved in assessment practices need to acquire. Stiggins identified an assessment literate person as one who could discern between excellent and poor quality assessment and apply that knowledge to make inferences about learner's achievements; however, he did not specify characteristics of such competence. This early attempt was followed by another one in 1999, in which Stiggins presented seven competencies that teachers need to acquire. He explained that for teachers to be considered assessment literate, they need to connect assessments to clear purposes, clarify

achievement expectations, apply proper assessment methods, develop quality scoring criteria, avoid bias in assessment, communicate effectively about students' achievements, and finally use assessments as an instructional intervention. These early attempts echoed the standards offered by AFT, NCME, and NEA in 1990; however, they have not added further dimensions. Deluca & Klinger (2010: 419-420) gave another dimension to TAL; they defined it as "the understanding and appropriate use of assessment practices along with the knowledge of theoretical and philosophical underpinning in the measurement of students' learning." Their definition added a theoretical dimension to TAL definitions and necessity to acquire knowledge of learning theory.

Popham (2011: 267) defined TAL as "an individual understanding of the fundamental assessment concepts and procedures deemed likely to influence educational decisions". He explained that TAL includes a set of knowledge and skills that teachers need to acquire. Teachers should be able to make inferences based on both learners' overt performance and covert skills and knowledge. This view implies that teachers should know how to construct both direct and indirect assessments. Popham added that teachers should be aware of issues that might threaten assessment such as reliability, validity, and fairness to avoid making invalid, unreliable biased inferences. Popham also focused on teachers' ability to construct selected-response and constructed-response test items in addition to designing and scoring rubrics. He paid special attention in his explanation to teachers' ability to construct performance assessments such as portfolios, in addition to being skilled in designing suitable rubrics to score them. According to Popham, assessment competence for teachers should not be limited to designing classroom-assessments only; it should involve interpreting students' performance on large-scale standardized achievement and aptitude assessments and scoring them, in addition to evaluating the appropriateness of such accountability assessments.

In the previous definitions, LTAL definitions were rooted in general education definitions of TAL, and no specific attention was given to it (Stabler-Havener, 2018). Different researchers tried to define LTAL; however, the reviewed literature revealed limited consensus on the definition and the set of knowledge and skills that constitute LTAL (Walters, 2010). Inbar-Lourie (2008a)

defined LTAL as a language teacher's understanding and knowledge about language learning theories, classroom assessment practices and the effective use of this knowledge to gauge and improve students' learning by employing various assessment method and strategies. Davies (2008) stated that LTAL comprises three fundamental components, skills, knowledge, and principles. Skills referred to teachers' expertise in conducting various assessments, assessment writing, and evaluation. Knowledge represented their background knowledge about assessment, language learning theories, and classroom pedagogy. Finally, principles signified their conceptual and practical cognizance of assessment qualities, such as validity, reliability, washback, practicality, interactiveness and authenticity.

Fulcher (2012: 125) gave a more expanded definition of LTAL that attempted to include a socio-cultural dimension; he defined LTAL as:

The knowledge, skills, and abilities required to design, develop, maintain, or evaluate large scale standardized and classroom-based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles, and concepts within historical, social, political and philosophical frameworks in order to understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, instructions, and individuals. (Fulcher, 2012:125)

Fulcher believes that LTAL is not merely limited to acquiring specific knowledge, skills or competencies; it is formed on three levels. Teachers should first acquire theoretical knowledge and skills that enable them to practice assessment tasks in the form of designing and evaluating both classroom assessment and standardized ones; which he called the practice of language testing level. This level is guided by a higher level, which he called the principles level, in which teachers need to apply ethical consideration and codes of practice to their assessment practices. Teachers' knowledge, skills, principles, and ethics are informed by a third and higher level, in which teachers pay attention to contextual considerations that control their assessment practices, such as

reasons for and impact of assessments practices in addition to considering historical, social, and political constraints that control their contexts.

Although Fulcher's definition involved a socio-cultural dimension to LTAL, out of all the previously mentioned definitions, I believe that the definition presented by Willis et al. (2013:242) is more detailed, inclusive, and closest to the philosophy of the current study. Although Willis' definition is focused on TAL in general and not specifically on LTAL, it emphasises that reality is socially constructed, subjective, changeable, and multiple. It applies a socio-cultural lens to the explanation of TAL. TAL is seen as a social practice where negotiation and cultural knowledge play essential roles in its formation along with other fundamental components such as knowledge and skills; moreover, it views TAL as a dynamic entity that is continuously developed. Willis et al. (2013) defined TAL as:

A dynamic context-dependent social practice that involves teacher articulating and negotiating classroom and cultural knowledge with one another and with learners, in the initiation, development and practice of assessment to achieve the learning goal of students. (Willis et al., 2013:242)

According to Willis et al. (2013), different assessment literacies are needed for different communities of practice, as they are different social groups (Hipwell & Klenowski, 2011). This view of multiple literacies is rooted in Bernstein's (1999) analysis of AL. He believed that knowledge is acquired and transmitted in two ways, vertical and horizontal. The former refers to the official or schooled knowledge, which is coherent, explicit, systematically principled, hierarchically organized, and takes the form of specialized language. The latter is local common sense knowledge, which is oral, context-dependent, specific, and multi-layered learning; it requires an oral transition in a social interactional relationship; however, this competence is tied to its social context and will not transfer to other contexts.

Accordingly, a definition of TAL would go beyond the traditional view of skills, knowledge, and cognition to culturally responsive practices through which teachers open up opportunities to negotiate their assessment identity/literacy and

practices within the learning community (Adie, 2010; Klenowski, 2009; Willis, 2010). Therefore, TAL would involve this vertical knowledge learned through official educational channels, whether theoretically or practically, and a horizontal knowledge, which is acquired through interaction with contextual practices and interaction with members of the same community in the form of reflection on oneself and others' actions and practices. However, this view assumes that TAL requires continuous learning about new curriculum and assessment policies. This dynamically formed literacy is formed as a result of continuous development that happens due to continuous exposure to new trends in assessment and new contextual assessment practices and experience. Teachers are required to develop new repertoires continuously. It is not just a particular fixed set of capabilities.

## **2.2 Research on Teachers' Assessment Literacy (TAL)**

The definitions mentioned above were the result of different research work conducted on TAL. The first stream of research focused on the knowledge base and skills of AL. This stream was followed by a second stream of research that focused on designing measurement tools to measure TAL to judge them as assessment literate or illiterate; however, it ignored contextual experience and knowledge that teachers might have acquired during their daily APs and being in contact with micro and macro contextual factors. The findings of the second stream of research revealed gaps in TAL; this tempted different researchers to explore assessment courses meant for pre-service teachers' preparation, in addition to exploring the effectiveness of teacher training programs for in-service teachers. The results of this stream of research revealed teachers' and researchers' dissatisfaction with assessment courses and training programmes, which teachers were exposed to before they start their professional career or during their teaching career. This dissatisfaction raised questions about teacher assessment education and preparation, which in turn led to the fourth stream of research, which focused on teachers' needs regarding AL preparation and development. The problem with this stream of research is that it depended on collecting data on teachers' assessment needs through checklists and questionnaires, which were based on pre-determined knowledge and skills; it ignored other context-based assessment components that could have been



included in these tools. This fourth stream paved the way to the fifth stream of research that focused on exploring teachers' assessment literacy in practice. The following section will explore these research streams thoroughly.

### **2.2.1 Research on Components of Assessment Literacy**

The first line of research into TAL began with attempts to describe the appropriate content and standards of knowledge and skills needed by teachers in order to be considered assessment literate. The goal of this stream of research was to establish a "knowledge base" for teachers (Xu & Brown, 2016). AL standards documents were created to shape and guide teachers' APs (Deluca, McEwan, & Luhanga, 2016). These standards were designed for teachers in general; they were not specifically addressed to language teachers; however, language teachers were affected by them in terms of required competencies. Two significant bodies produced the standards:

#### ***Governments and Research-Based Assessment Organizations***

One of the earliest attempts to address AL was initiated by AFT, NCME, and NEA in 1990; however, the term "assessment literacy" was not in use at that time (Fulcher, 2012). AFT, NCME, & NEA (1990) listed several assessment competencies that teachers need to acquire if they are involved in different APs, whether large standardized ones or classroom-based ones. These competencies included: (1) Selecting assessments methods relevant to instructional purposes. (2) Developing assessment methods appropriate for instructional decisions. (3) Administering, scoring and evaluating assessment methods, in addition to interpreting the results of both externally produced and teacher-produced assessment methods. (4) Using assessment results in decision making related to improving instructions, students, school, and curriculum. (5) Developing proper grading procedures. (6) Communicating assessment results with students, parents, and administration. (7) Recognizing illegal and unethical assessment practices, methods and information (AFT et al. 1990).

This initial attempt to list AL standards was followed by further attempts from different research bodies and individual researchers to fulfil the same aim (Deluca, McEwan, & Luhanga, 2016). Some of these attempts are listed in table 2.1.

<b>Research Body/Researcher</b>	<b>Literacy Standard</b>	<b>Year of Publication</b>
<i>Joint Advisory Committee (JAC)</i>	The principles for fair students' assessment practices for education in Canada	1993
<i>National Council on Measurement in Education (NCME), USA</i>	The code of professional responsibilities in measurement	1995
<i>The Australian Curriculum &amp; Assessments and Certificate Authorities (ACACA)</i>	Guidelines for assessment quality and equity	1995
<i>The American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME), USA.</i>	Standards for educational and psychological testing	1999
<i>Assessment Reform Group (ARG), UK.</i>	Changing assessment practices: process, principles and standards	2008
<i>Susan M. Brookhart</i>	A set of educational assessment knowledge and skills for teachers	(2011)
<i>The Association for Educational Assessment (AEA), Europe</i>	The European framework of standards for education assessment	2012
<i>The Joint Committee for Standards on Educational Evaluation (JCSEE)</i>	Classroom assessment standards	2015

*Table 2.1 Governments and Research-Based Assessment Organizations Which Specified Assessment Competencies*

### **Teacher Accreditation - and Certification – Based Organizations**

The following are professional attempts adopted by teacher accreditation and certification organizations to set standards for TAL:

<b>Accreditation Organization/Certification</b>	<b>Literacy Standard</b>	<b>Year of Publication</b>
---	--------------------------	----------------------------

<i>The National Board for Professional Teaching Standards (NBPTS), USA.</i>	A list of what teachers should know and do	2001
<i>The National Council for the Accreditation of Teacher Education (NCATE), USA.</i>	Profession standards for the accreditation of teachers' preparation	2008
<i>New Zealand Teacher Council</i>	Graduating teacher standards	2008
<i>The Interstate Teachers Assessment and Support Consortium (INTASC)</i>	The model core of teaching standards	2011
<i>Department of Education, Australia.</i>	The Australian professional standards for teachers	2012
<i>Department for Education, United Kingdom (UK).</i>	Teachers' standards	2012

*Table 2.2 Professional Attempts Adopted by Teacher Accreditation and Certification Organizations to Set Standards for TAL*

Deluca, McEwan, & Luhanga (2016) analysed the proposed standards of all previous 15 attempts, whether research-based or teacher accreditation-based using thematic coding procedures adopted from Patton (2002). They came up with eight common themes that enabled them to report on the content of the 15 attempts. The themes were: (1) Assessment Purposes: choosing the appropriate form of assessment based on clearly stated instructional goals. (2) Assessment Process: constructing, administering, scoring assessments, and interpreting results to facilitate instructional decision making. (3) Communication of Assessment Results: communicating assessment purpose, process, and results to stakeholders. (4) Assessment Fairness: creating fair assessment conditions for all learners with sensitivity to learners' diversity and exceptional learners. (5) Assessment Ethics: disclosing accurate information, protecting rights and privacy, minimalizing biases, and complying with standards. (6) Measurement Theories: understanding the psychometric properties of assessments (reliability, validity, norms, and standards). (7) Assessments for learning: using formative assessments during instructions to guide teachers' practices and students' practices (8) Education and Support for Teachers: educating and supporting teachers' assessment competency. These eight components of AL did not appear

in all standards. Deluca, McEwan, & Luhanga (2016) reported the findings of their thematic analysis as follows:

1990-1999: Early documents focused on teachers' selection and use of assessments, mainly summative and standardized assessments with an emphasis on developing teachers' psychometric understandings. This period focused on developing TAL about assessment purposes, process, result communication, and assessment fairness. This kind of literacy reflected the demands of the accountability movement that dominated during this period. Therefore, assessment standards emphasised and incorporated learning measurement theories, assessment ethics, reliability, validity, norms, protecting learners' rights and privacy. The only document in this period that supported teachers' assessment education as one component of assessment standards is the NCME issued in 1995.

2002-2009: Knowledge of assessment purposes, process, result communication strategies, and assessment fairness remained central as the main component of TAL standards; however, new themes emerged. Assessment for learning theme appeared in standards published in 2001 and 2008. Support for teachers' education appeared as a central component of AL standards in the NCATE in 2008. This period reflected the attention paid to integrating assessment practices with pedagogy and use of assessment to guide teaching and learning.

2010-Present: This period witnessed the inclusion of all previously mentioned themes as central components of TAL. The European standards document is the only exception; as it did not include assessment for learning and assessment education support for teachers as components of the document. Standards documents that appeared in this period (Brookhart, 2011 & JCSEE, 2015) included a critical component as one of its standards, a component that was introduced only during this period.

Out of all the previously explored standards, I believe Brookhart's (2011) standards stood out, as they reflect recent changes in educational pedagogy. Brookhart mentioned that although formative uses of assessment appeared in

AFT standards, their focus was only on teachers' formative use of assessment information in decision making not on students' formative use of assessment information. New language learning theories emphasise how students learn and self-regulate their learning based on formative assessment information they receive from teachers, peers, and their interpretation that enable them to make effective decisions about their learning (Black & William, 2009). Student involvement in using assessment information requires an extension in teachers' assessment knowledge and skills, as this will require teachers to be skilled in clarifying and sharing learning intentions and criteria for success with their learners and providing meaningful opportunities for students to take actions based on the assessment information.

Moreover, Brookhart argued that the standard-based reform movement required additional competences than what the standards mentioned above offered. Schools are no longer judged based on physical resources, teachers' qualification, and human resources only; educational outcome and student achievement were the primary sources of judging schools and teachers and making them accountable for such outcomes. Consequently, teachers will need knowledge and skills on how to use these accountability standards and adapt their daily instructions and assessment practices to contribute to standard-based accountability and achievement. Accordingly, Brookhart proposed a set of assessment competences based on those innovative requirements that teachers need while performing assessment tasks:

(1) Teachers should understand learning in the content area they teach. Content knowledge and pedagogical content knowledge are essential for teachers before performing assessment tasks.

(2) Teachers should be able to articulate clear learning intentions that align with both content and curriculum goals. Teachers should be able to write the goals and objectives of their instruction and consequently, their assessments in precise ways.

(3) Teachers should be able to communicate the criteria for success, especially with formative assessments (William, 2010) to learners and colleagues. Teachers

can communicate criteria for success through different communication modes like telling, showing, and having learners discover (Moss & Brookhart, 2009).

(4) Teachers should understand the purpose and uses of available assessment options and be skilled at using them. Assessment can be meant for formative purpose to obtain information on both teachers' and learners' performances; it could be held for summative purpose to use the information for reporting or placement decisions. Teachers must know what kind of test items and performance tasks will draw out the knowledge and skills they intend to measure (target method match). Teachers need to understand the concept of validity, including reliability. Teachers should also understand the issues of fairness and accessibility.

(5) Teachers should have the skills to analyse classroom question, test items, and performance assessment tasks to assess the required knowledge and thinking skills. They should understand the kinds of cognitions taxonomies like Bloom's and Webb's to know the thinking skills required for different test items or tasks.

(6) Teacher should have the skills to provide effective useful feedback on students' work, whether descriptive, elaborative, or specific.

(7) Teachers should be able to construct scoring schemes that quantify learners' performance into useful information for decisions, which in turn should lead to improved learning, growth, and development.

(8) Teachers should be able to administer external assessments and interpret their results for decisions about learners. Such knowledge or skills will help them to improve students' learning and reduce some of the pressure that comes with current accountability policies.

(9) Teachers should be able to articulate their interpretation of assessment result and their justification for decisions taken accordingly. They should explain to students the meaning of the results and means of improving them. They need to

be skilled in communicating their interpretations. They need to participate in educational reform committees.

(10) Teachers should be able to help learners use assessment information to make sound educational decisions.

(11) Teachers should understand how to perform legal and ethical responsibilities while performing all assessment steps and process. (Kanjee & Mthembu, 2015:146).

Reviewing assessment standards/competencies, it is evident that setting goals for teachers and teachers' educators in the form of guidelines and standards are fundamental; however, their existence alone does not guarantee improvements in teachers' assessment education (Xu & Brown, 2016). This knowledge base is not a static one; it is dynamic due to the emergence of new theories, research findings, and socio-cultural changes. Therefore, teachers should keep themselves aware of new trends in assessments to keep their knowledge base up to date and to be able to experiment with new ideas, which might be solutions to challenges they encounter in their assessment contexts.

### **2.2.2 Research on Measurements of Assessments Literacy**

As a result of using assessment standards as guidelines for teachers' assessment practices, instruments were developed to investigate teachers AL levels to judge their strength and weakness in the prescribed standards or competencies (Xu & Brown, 2016). The first six instruments were based on AFT, and the last two were based on other frameworks. The below table summarises eight measurement instruments used to assess LTAL:

<b>Measurement Instrument</b>	<b>Developed By</b>	<b>Function</b>
<i>Assessment in Vocational Classroom Questionnaire</i>	Kershaw IV (1993)	To measure in-service teachers' level of competence in assessment activities using 26 items, each with a 5-point scale.

<i>Teachers Assessment Literacy Questionnaire (TALQ)</i>	Plake et al. (1993)	To measure in-service teachers' assessment literacy in the seven standards articulated in 1990. It consisted of 3 items, five for each standard.
<i>Assessment Practices Inventory (API)</i>	Zhang & Burry-stock (1997)	To measure in-service teachers' perceptions of their assessment skills using 67 items, each with a 7-point scale that ranged from not confident to very confident. Items were grouped into seven groups.
<i>The Measurement Literacy Questionnaire</i>	Daniel & King (1998)	It was based on assessment literature (Kubiszyn & Borch, 1996; Popham, 1995).
<i>Assessment Literacy Inventory (ALT)</i>	Campbell et al. (2002)	It was based on the TALQ one; however, it was administrated to measure pre-service teachers' assessment literacy.
<i>Classroom Assessment Literacy Inventory (CALI)</i>	Mertler (2003)	It was developed based on TALQ to measure both pre-service and in-service teachers' assessment literacy.
<i>Revised Assessment Literacy Inventory (RALT)</i>	Mertler & Campbell (2005)	This instrument was developed in response to calls to revise TALQ to measure pre-service teachers' assessment literacy.
<i>Assessment Self-Confidence Survey</i>	Jarr (2012)	To measure teachers' self-confidence with assessment-related practices. It was developed based on Bandura's (2006) guidelines. It consists of 15 Likert-type items.

*Table 2.3 Eight LTAL Measurement Instruments*

Different empirical studies used the instruments mentioned above to measure LTAL (Alkharusi, Kazem, & AIMusawai, 2011; Maclellan, 2004); however, the consensus was that teacher assessment knowledge was generally inadequate relative to standards expectation (Deluca, McEwan, & Luhanga (2016). Studies that investigated the level of AL among in-service teachers



revealed that language instructors showed low or moderate levels of AL (Kiomrs, Abdolmehdi, & Naser, 2011; Leoph, Chamy, & Chan, 2015; Talib, Kamsah, Ghafar, Zakaria, & Naim, 2013; Xu & Brown, 2017).

Analysing the items of the previously mentioned instruments, Deluca, McEwan, & Luhanga (2016) found that assessment process was the most commonly represented theme in evaluating LTAL; assessments' results communication, assessment ethics, and assessment purposes themes were often equally represented in measurement instruments of LTAL. Assessment fairness and assessment for learning themes were represented in two instruments only. None of the instruments included themes for measuring assessment education and support for teachers.

Deluca, McEwan, & Luhanga (2016) recommended that AL measurement instruments should reflect the complexity of AL constructs. Utilizing new added standards as a guiding framework for constructing measures would promote the validity of these instruments (Kane, 2006). Deluca et al. suggested considering components, such as assessment education and support for teachers when constructing AL measures. They recommended seeking information on teachers' preferences, experiences, and perceived effectiveness of assessment education. They argued for specifying the population of the instrument (pre-service/in-service), as they have different learning needs. They supported investigating teachers' actual APs, as this will establish the value and validity of the used AL measurement instruments. Finally, they emphasised using measurement instruments in provoking positive learning consequences for teachers.

Although using instruments for measuring teachers' mastery of knowledge base and principles of assessment gave an estimate of TAL, it showed inadequacy in teacher assessment preparation (Brown & Xu, 2016). This finding led to the third stream of research in AL that focused on exploring teachers' assessment education and its impact on assessment practices.

### **2.2.3 Research on Language Assessment Education**

Different studies were conducted in the field of teachers' assessment education on different levels, as presented below.

### **2.2.3.1 Assessment Courses**

The first group of studies focused mainly on exploring and assessing the quality of LA courses for pre-service teachers (Greenberg & Walsh, 2012). These studies were conducted on course content and its inclusion. They found that many teachers' preparation courses lack an assessment component. Bailey & Brown (1995) and Brown & Bailey (2008) looked at the content of language testing programs and the textbooks proposed to pre-service language teachers; they found that testing courses covered basic formal APs as well as essential descriptive statistics, validity, and reliability. They detected a lack of preparation for formative assessments design and scoring. Pre-service teachers were trained on test critique, item writing, interpreting test scores, scoring tests, and administering tests. In a study conducted by Deluca & McEwen (2007) on pre-service teacher education in Canada, they found that 3 of 10 bachelor education programs in Ontario had a mandatory assessment course or module. Volante & Fazio (2007) explored AL of pre-service teachers in the same Canadian context, and they reported a low level of self-efficiency in each year of the program. In another study conducted by Jin (2010) in China, the inadequacy of pre-service assessment training was reported in the Chinese context. Lam (2015) found a similar finding in Hong Kong among pre-service teachers. Moreover, he found that some programs offered assessment courses as electives.

Similarly, research into in-service assessment education showed similar results. Vogt & Tsagari (2014) found in their study that explored in-service language teachers' assessment education in Europe that only some aspects of their AL are developed; they compensate for this insufficiency of training by learning on the job or using teaching materials that are ready for use. Teachers expressed need to receive assessment training.

### **2.2.3.2 Assessment Training Programs, Tools & Recourses**

Another group of studies explored the effectiveness of practical training opportunities on both pre-and in-service teachers. As per explored studies, for training programs to be effective with pre-service teachers, several conditions should be met: 1. Assessment education should take various forms and integrate different stakeholders' perspectives (Deluca, 2012; Hill, Ell, Grudnoff, & Limbrick, 2014; Mertler, 2009). 2. AL should become part of teacher accreditation and

certification (Sato, Wei &, Darling-Hammond, 2008). 3. Mentors should attend to pre-service teachers' prior beliefs on assessment (Graham, 2005). 4. The training content should be localized. (Lam, 2015; Leahy & William, 2012).

For in-service teachers, some studies found that theoretical professional training in language assessment improved in-service teachers' assessment literacy (Mahapatra, 2016; Montee, Bach, Donovan, & Thompson, 2013; Nier, Donovan, & Malone, 2013; Walters, 2010). In-service teachers who have limited opportunity to learn through formal training channels may need to learn from online learning resources (Fan, Wang, & Wang, 2011), seek support from within the workplace (Lukin, Bandalos, Eckhout, & Mickelson, 2004), and utilize daily classroom practices as sites for their professional development by implementing assessment for learning (Smith, 2011). However, seeking these resources does not deny the need for sustainable assessment training.

Different empirical studies reported a positive relation between assessment training and teachers AL (Levy-Vered & Alhija, 2015; Quilter & Gallini, 2000); however, the fact that some training courses for pre-service or in-service teachers can be pre-determined enforced ones raises questions regarding teachers' conceptions about them. Some studies found that teachers might resist training because the amount and type of training in assessment was independent of their beliefs about assessments (Brown, 2008a). Teachers have their deeply-rooted conceptions of assessment, which may constitute a significant barrier to developing teacher AL (Quilter & Gallini, 2000). Assessment education needs to encompass both technical knowledge of assessment and more consciousness arousing components that prompt teachers to examine their conceptions (Deluca & Lam, 2014; Hill, Cowie, Gilmore, & Smith, 2010). This concept paved the way for more studies in the area of teacher training needs.

### **2.2.3.3 *Teachers' Assessment Training Needs***

Another group of studies investigated the assessment training needs of language teachers either to inform writing textbooks on AL or to act as the knowledge base for constructing professional development tools that aim to enhance TAL, such as workshops, manuals, and handbooks. Hasselgreen, Carlsen, & Helness (2004) and Huhta, Hirvala, & Banerjee (2005) conducted

quantitative studies to uncover the assessment training needs of teachers across Europe using a survey. They found that teachers need training in portfolio assessment, preparing classroom tests, peer and self-assessments, interpreting test results, continuous assessments, giving feedback on work, validity, reliability, statistic, item writing, item analysis, and interviewing. A similar study was conducted by Vogt & Tsagari (2014) in 7 European countries with the same aim; however, they used mixed-methodology (survey and interviews) to elicit TAL needs from 878 pre-service and in-service language teachers. The findings revealed insufficiency in teachers' assessment preparations and their needs to more assessment training, especially in unfamiliar forms of assessment such as portfolio, self, and peer assessments. These findings echo the results of the previous studies. Fulcher (2012) conducted a study to elicit assessment needs of 278 language teachers to inform the design of new teaching material on language testing and online resources that could support program delivery. He collected both quantitative and qualitative data through questionnaires and semi-structured interviews. His findings revealed that teachers need an approach to AL that integrates knowledge, skills and principles in a procedural text that attempts to balance what is required for both classroom and formative assessments.

Despite the importance of such studies and the amount of needs teachers expressed, Deluca & Klinger (2010) mentioned that in many cases teachers expressed a need for "everything" rather than being specific about their needs; other studies revealed that teachers lacked confidence in their APs (Volante & Fazio, 2007). These findings mean that teachers' perception of their assessment training needs may not be what they precisely want (Xu & brown, 2016). These results suggest that a contextual understanding of AL may provide a more useful approach to teacher assessment education because it provides wider learning opportunities of assessment knowledge in the field through reflection and collaboration about classroom experiences (Howley, Howley, Henning, Gilla, & Weade, 2013).

#### **2.2.4 Research on Understanding/Developing Assessment Literacy in Practice**

The underpinning assumption for research conducted on teachers' education was that TAL would improve if pre-and in-service programs prepare teachers sufficiently, however, another body of research started to study contextual factors and its impact on shaping TAL. Xu & Brown (2016) believed that these contextual factors inhibit the effectiveness of teachers' AL. On the national level, Forsberg & Wermke, (2012) argued that national policies adopted by countries influence TAL by shaping their professional development activities and autonomy; Gu (2014) added that imposed policies dictate curriculum standards, textbook use, and large-scale tests, which in turn have an impact on teachers' practices and accordingly assessment practices and beliefs. On the institution level, structural conditions like power relations in the workplace shape teachers' practical knowledge of assessment (Xu & Liu, 2009). On a personal level, teachers' awareness of and actions to construct their identity as assessors also matter (Adie, 2013; Cowie, Cooper, & Ussher, 2014; Scarino, 2013).

The result of this body of research, which is informed by a socio-cultural view of learning, reinforced a more integrative approach towards exploring and developing TAL; one that is shaped and informed by teachers' assessment conceptions, practices, and beliefs of assessments situated in their specific contexts rather than on the mastery of assessment knowledge, skills and principles (Frey & Fisher, 2009; Wyatt-Smith, Klenowski, & Gunn, 2010). In this sense, teachers are the main drivers of their own AL development; and consequently will make full use of classroom-based assessments as sites for learning (Gottheiner & Siegel, 2012), engage in regular dialogues and collaboration with colleagues (Wyatt-Smith et al., 2010), and participate in assessment activities within communities of practices (Willis et al., 2013). Scarino (2013) argued that teachers placed in complex contexts have to make professional decisions about assessments in response to various factors. They need to balance the demands of external factors with their own beliefs and values (McMillan, 2003). Solutions to improve teacher AL are by no mean universal but rather contextual (Xu & Brown, 2016).

## **2.3 Language Assessment Literacy Models/Frameworks**

In correspondence to TAL research stages mentioned above, researchers focused on designing assessment models or frameworks to act as the backbone for training programs of AL or theoretical reference through which TAL would be explored in different contexts. Some of these models or frameworks focused mainly on the knowledge base, skills and principles of AL, as they were affected by early studies; however, recent models presented a more holistic view of AL that depends on a combination of cognitive traits, effective system, and socio-cultural and institutional inference. They "pinpoint the complexity of LAL" (Xu & Brown, 2016, p., 155).

### **2.3.1 A Five Component Model (Brindley, 2001)**

Brindley was the first language tester to present a model of AL; he believed that different stakeholders need to acquire different levels of assessment knowledge according to the nature and extent of their involvement in the assessment process. He proposed five units that form AL, two core units and three optional units. The first core unit involves knowledge about the rationale for the assessment. Teachers need to be aware of the social characteristics, in which assessments take place because contexts differ in how they value and perceive assessments. Ethical awareness is also crucial; it would be unethical to base decision on poorly designed tests that are not valid or reliable. Teachers awareness of political policies and practices that might discriminate, classify, exclude individuals will help teachers avoid bias when constructing assessments. The other core unit includes knowledge about language proficiency (theoretical models of language models). As for the three optional units, one detailed knowledge of developing and evaluating language tests in addition to knowledge of statistical analysis; another detailed knowledge about criterion-referenced tests and techniques associated with alternative assessments, and the final unit involved knowledge about strategies for applying and putting into practice issues raised in the previous models.

Inbar-Lourie (2008a) reviewed Brindley' model and modified his themes; she was against the idea of having some of the units optional, as she believed that assessors need all units to perform their assessment tasks. She proposed a

model of LAL that involved what she called core knowledge of LAL; she argued that all participants in the assessment process should obtain this core knowledge. However, the focus and intensity would vary and depend on the target audience. She explained: "Becoming assessment literate requires the attainment of a "toolbox" of competencies, some practical and some theoretical on "why", "what" and "how" to go about constructing a variety of assessment procedures" (Inbar-Lourie, 2008a:389). "Why" provides the reason for the assessment, which is equivalent to core unit one in Brindley's. "What" describes the characteristics of the language ability to be assessed, which is equivalent to core unit two in Brindley's. "How" specifies the methods of assessment (equivalent to the optional units). She explained that this core knowledge would enable language assessors to "gain an initial footing in the field and speak the language of assessment; additional competence can be developed, refined and elaborated for specific purposes" (Inbar-Lourie, 2008a:396). Similar to the Brindley's model, Inbar-Lourie's is seen as a magic package that teachers can have access to and use in every context; a view that does not consider uncertainties, diversities and challenges that teachers see in their daily professional life.

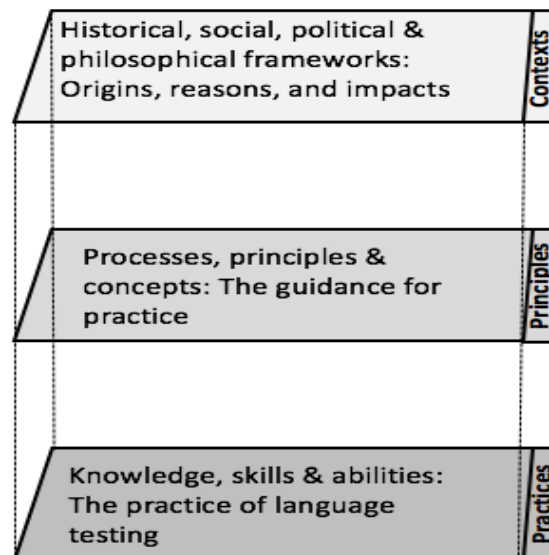
### **2.3.2 A Skill, Knowledge, and Principles Model (Davies, 2008)**

Davies presented a model that divided LAL into knowledge, skills, and principles. For knowledge, Davies believed that language teachers should be provided with a foundation in measurement and language proficiency as well as contextual considerations. For skills, he believed that language teachers should be offered education in methodology and tools such as item writing, statistics, analysis, and reportage as well as software programs that assist teachers in their assessment tasks. For principles, he argued that teachers should be trained in the use, impact, fairness, and ethics of assessments. These three components remained constant in theoretical and research discussions about LAL; however, their boundaries have been questioned (Inbar-Lourie, 2013; Taylor, 2013).

### **2.3.3 A Practice, Principles, and Contexts Model (Fulcher, 2012)**

Fulcher constructed a diagram that implies a hierarchy in LAL acquisition, with practices (knowledge, skill, and abilities involved in language testing) as the base foundation of LAL. The next level is principles, which involve principles, processes, and concepts that guide foundational practices. The top-level is

contexts, which involves historical, social, and political philosophical frameworks in which the practices and principles are placed. I believe that what makes Fulcher's model different from the previous model is his inclusion of a social component to LTAL. Fulcher placed the theoretical and practical assessment knowledge proposed by the previous models as the base of LTAL. This knowledge is guided by a social framework that shapes, guides, and informs teachers on what could work in assessment practices.



*Figure2.1 Language Assessment Literacy Model (Fulcher, 2012, p. 126)*

#### **2.3.4 LAL Stakeholders Profile Model (Taylor, 2013)**

Taylor (2013) proposed a model of LAL that placed different stakeholders involved in language assessments at different levels. Not everyone needs to know or be able to do everything to the same level. She believed that some stakeholder groups should be close to the heart of assessment; other stakeholder groups might place themselves some distance away from the core or closer to the periphery. Researchers and test makers were at the core of the figure. Such groups should receive comprehensive training in measurement theory, technical know-how and ethical principles. Language teachers and course instructors were placed at an intermediary level, and policymakers and the general public were on peripheral levels of LAL. "The different stakeholder groups were described against eight dimensions (Knowledge of theory, technical knowledge, principles and concepts, language pedagogy, socio-cultural values, local practices, personal beliefs & attitudes, and scores & decision making" (Taylor, 2013: 410).



Taylor created different profiles for test writers, classroom teachers, university administrators, and professional language tester. She gave values to each dimension (from 0 to 4) based on Pill & Harding (2013), who classified LAL proficiency levels into (1) Illiteracy (ignorance). (2) Normal literacy (basic understanding with some misconceptions). (3) Functional Literacy (Sound basic understanding). (4) Procedural and Conceptual Literacy (understanding and practice of central concepts). (5) Multidimensional literacy (extended knowledge beyond ordinary concepts). Figure 2.2 describes LAL profile for classroom teachers, in which Taylor believed that classroom teachers should have (1) multidimensional literacy in language pedagogy, (2) procedural and conceptual literacy in technical skills, local practices, socio-cultural values, personal beliefs, and (3) normal literacy in the knowledge of theory, scores and decision making, and principles and concepts.

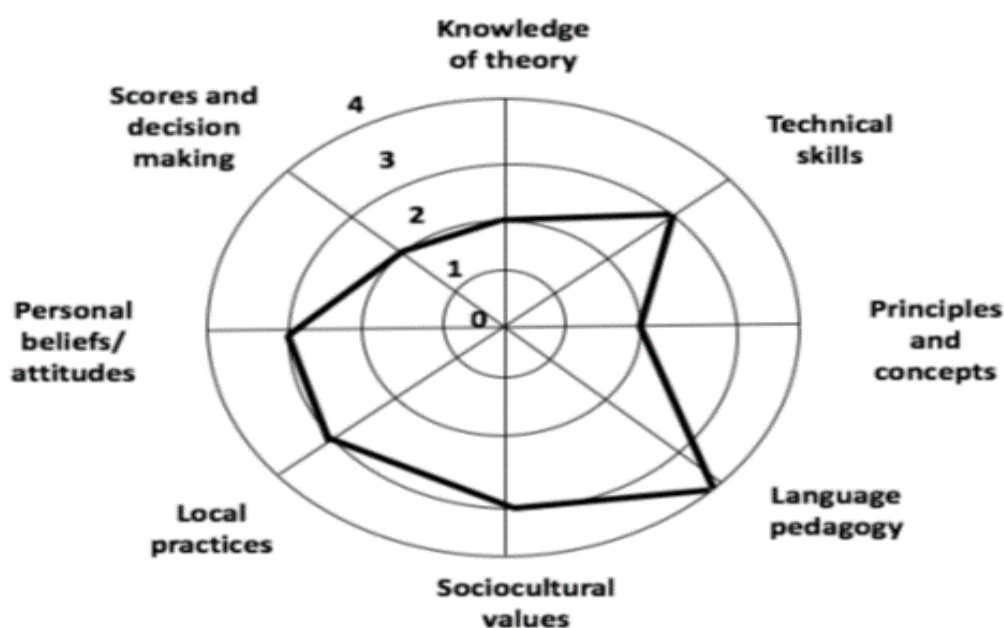


Figure 2.2 LAL Profile for Classroom Teachers (Taylor, 2013, p. 410)

In Taylor's profiles, none of the stakeholder groups should be ignorant in any of the assessment dimensions; every group should be competent in all dimensions, but with different degrees according to their involvement level and required assessment tasks. Taylor's inclusion of personal beliefs added a third dimension to TAL; this echoed Scarino's (2013) beliefs about the impact of

teachers' beliefs on their willingness to adopt new educational policies and their assessment awareness in general. Taylor emphasised the fact that LAL is not necessarily a value-free concept; however, her model is criticized for being speculative; some researchers argue that classroom teachers require more advance literacy (Harding & Kremmel, 2016).

### **2.3.5 Teacher Assessment Literacy in Practice (TALiP) Model (Xu & Brown, 2016)**

Xu & Brown (2016) proposed a conceptual framework of Teacher Assessment Literacy in Practice (TALiP), which consists of six components: (1) knowledge base, (2) teacher conceptions of assessment, (3) institutional and socio-cultural contexts, (4) TALiP the core concept of the framework, (5) teacher learning, and (6) teacher identity (re)construction as assessors.

As per the proposed framework, Xu & Brown (2016) recommended taking into consideration the suggested six components while exploring or developing TAL. On the first level, exploring/developing assessment knowledge base is fundamental. This level includes basic assessment theoretical knowledge, skills, and principles without which teachers cannot engage with assessment at a deeper level. Xu & Brown (2016) compiled all assessment knowledge, skills, and principles proposed by previous AL models and proposed them as "knowledge base" of assessment. This "knowledge base" of assessment is derived from the AFT, NCME, & NEA, (1990) standards and recent updates (Brookhart, 2011; JCSEE, 2015), in addition to some contemporary assessment textbooks for teachers (Airasian & Russell, 2008; Brown, Irving, & Keegan, 2014; McMillan, 2001; Nitko & Brookhart, 2011). Xu & Brown believe that researchers cannot neglect this fundamental knowledge base if they are exploring TAL or designing models of AL development. Although they believe the knowledge base is a necessary component of TALiP, they argue that it is not a sufficient element of TAL.

On the second level, Xu & Brown believe that teachers' conceptual understanding of assessment needs to be paid attention to while exploring/developing TAL. This level is a personal conception of how assessment should be. This conception is formed as a result of tensions between theoretical

knowledge teachers were exposed to during their formal/non-formal assessment preparation and their own beliefs/views/emotions about assessment that were formed by their experience of assessments (Barnes, Fives, & Dacey, 2015; Fives & Buehl, 2012). There should be opportunities to re-examine these conceptions, mostly that they are neglected in AL research (Deluca & Lam, 2014; Hill, Cowie, Gilmore, & Smith, 2010). These deeply-rooted conceptions can sometimes act as barriers to any attempts of AL development if they contradict with the content of the proposed development opportunities, as teachers/assessors perceive and conceptualise assessment knowledge that is congruent with their beliefs/values/emotions and reject those that are not (Gunn & Gilmore, 2014; Quilter & Gallini, 2000).

Therefore, exploring teachers' conceptions about assessment knowledge base is crucial because it provides insights on what teachers believe is true or false regarding the proposed knowledge base and shows their emotional inclinations towards different aspects of assessment (Crossman, 2007). Exploring these conceptions can filter what is needed or not needed from the broad knowledge base and most important what could be implemented or used in their assessment contexts depending on their assessment conditions, roles, and purpose. Professional development efforts, which ignore teachers' emotional and conceptual views about the proposed knowledge base, are likely to be less successful (Brown, 2008a; DeLuca, Chavez, & Cao, 2013). Different studies have acknowledged the role teachers' conceptions play in shaping AL (Brown, 2008a; Levy-Vered & Alhija, 2015).

On the third level, Xu & Brown believe that attempts to explore/develop teachers' AL, which ignore the social micro-and macro- variables that influence teachers' APs through policies, norms, rules, and regulations are deemed incomplete and unrealistic (Scarino, 2013, p. 312). They explained that teachers' institutions set boundaries for their assessment practices in terms of what they should do (Gu, 2014). Moreover, teachers' APs are sometimes influenced by the needs/interests of various stakeholders, including school administrators, students, and colleagues (Wyatt-Smith et al., 2010; Xu & Liu, 2009). Therefore, exploring these contextual assessment boundaries/philosophies and teachers' beliefs about such adopted assessment philosophies, to which they are expected

to confront is essential because it can show the kind of incongruence that might arise between their conceptions and the boundaries imposed upon them within their context. The tighter the boundaries, the less space there is for professional autonomy. Tensions arise for teachers when they have less autonomy (Fleer, 2015; Forsberg & Wermke, 2012).

The fourth stage of the framework focuses on the kind of compromises or assessment decisions teachers take to balance the demands of external macro and micro-factors with their own beliefs and values (McMillan, 2003). Acceptable APs involve the art of compromise (Carless, 2011). If teachers do not question their conceptual assessment beliefs and their context-bound assessment philosophies, problematise them, and reflect on them, they would end up repeating traditional practices that are inconsistent with research evidence about useful APs. Therefore, engaging teachers in problem-solving situations, where they can critically reflect on context-bound philosophies and practices and try to find solutions to aroused problems can contrite to their AL conceptual development (Xu & Brown, 2016).

The last two levels of the proposed model are more concerned with teachers' role in their own AL development in their contexts. Teacher learning is a significant component in the TALiP framework; it can be done in their contexts in two ways: reflective practice (Schon, 1983) and participation in community activities (Westheimer, 2008). Reflection helps teachers understand the links between what they do and how they might improve their effectiveness (Cornish & Jenkins, 2012; DeLuca, 2012). Such reflective practise is critical to developing TALiP because it may open up a "change provoking disequilibrium" (Woolfolk, Hoy, & Davies, 2009: 645) in which teachers may realise the need to unlearn or relearn certain aspects. Participation in community activities engages teachers in professional conversations about their APs, offer opportunities to understand alternative thinking and practice of assessment, and allow them to defend their conceptions and negotiate their ideas with colleagues (Fleer, 2015; Leahy & William, 2012). Such engagement may lead teachers to make subsequent changes in their day-to-day APs and new compromises in the interest of student learning.

The last stage in the framework focuses on self-directed awareness of own assessment new role and identity, which were both formulated as a result of being engaged in such exploratory/developmental stages about own assessment concepts and practices. Teachers need to realise and notice the development that happened to them as a result of this professional engagement. This noticing phase will make them understand the value of assessment and how their educational decisions in assessment can make a positive difference to student learning. This awareness may empower them with autonomy, resources, and voice, and therefore help them in reclaiming their ownership of assessment practices, and open more opportunities, in which their AL can be recognised by assessment communities (Wyatt-Smith et al., 2010).

Therefore, it is crucial to involve assessors in research practices, where they can utter the results of conceptual development that happened to them because teachers who have a clearer sense of their own identity as assessors are likely to be better authors of their APs and to integrate others' perspectives into their values. They might likely change their practices, conceptions, and their traditional decision-making processes about student learning, reconsider alternative ways of practice, relearn and unlearn assessment knowledge, trial new assessment approaches or strategies, and consequently change their assessment practices to facilitate students' learning (Brown & Xu, 2016).

This conceptual framework of TALiP is illustrated in Figure 2.3. It views LTAL as multi-layered and multi-directional. The relation among the components of the TALiP model is an interrelated one. According to the TALiP model, TAL is built on three primary levels. On the first level, teachers are expected to master educational assessment knowledge, which includes the fundamental principles of assessments ("what", "why", and "how". The knowledge base consists of disciplinary knowledge and pedagogical content knowledge, knowledge of assessment purposes, content and methods, knowledge of grading, knowledge of feedback, knowledge of assessment interpretation and communication, knowledge about student involvement in assessment, and finally knowledge of assessment ethics.

On the second level, teachers are expected to develop an assessment awareness, which is their perception of how assessment should be. This awareness is developed as a result of two interactions, one between teachers' assessment knowledge base and their conception of it, and the other between teachers' assessment knowledge base and their contextual domains whether internal or external.

On the third level, a more advanced assessment awareness level is acquired through a self-directed approach and development of one's own identity as an assessor. This level will enable teachers to accommodate assessment policies and at the same time, reflect on their assessment practices to gain new insights. This model suggests different dynamic assessment literacies among various contexts; however, there are possible chances of having regularity and generalisations about how TALiP is developed, advanced and understood that could apply to all teachers across various contexts (Xu & Brown, 2016).

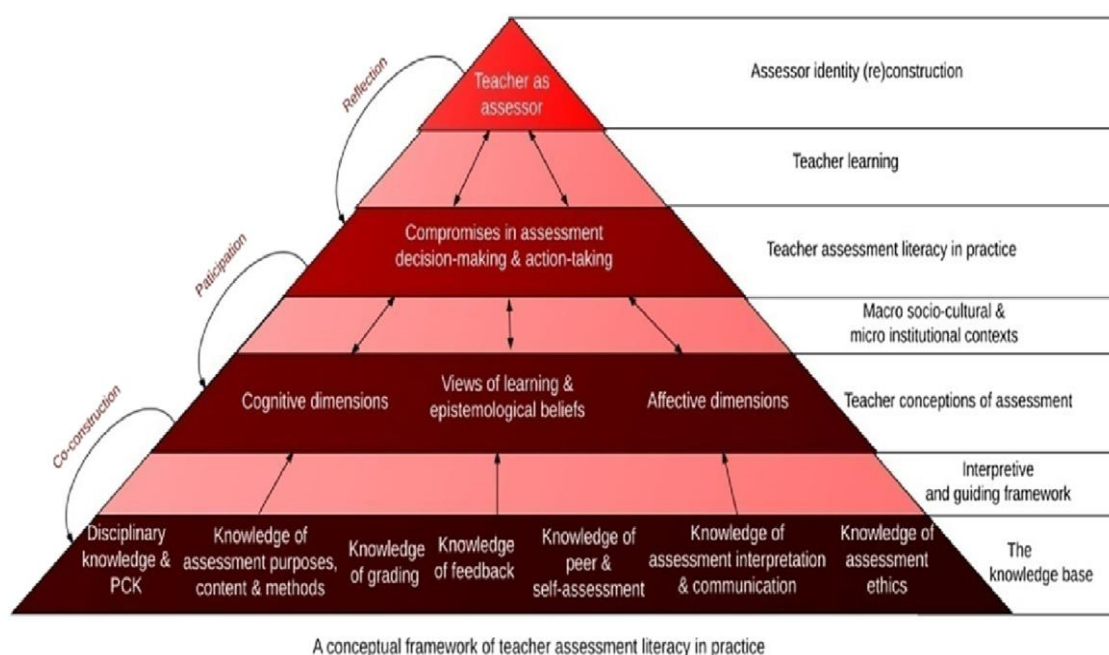


Figure 2.3. Conceptual Framework of TALIP Model (Xu & Brown, 2016: P. 155)

Based on the explored literature in section 2.3, it is concluded that TAL is a complex concept that cannot be explored apart from teachers' own beliefs and teaching contexts. I believe that awareness of formal systematic codified knowledge of assessment is necessary and essential for exploring or developing

LTAL. Without awareness of this knowledge base, there will be no standards through which assessment practices could be explored, evaluated or developed (Fulcher, 2012). However, exploring assessment knowledge base of teachers alone as an indicator of TAL is insufficient because they are decontextualised guidelines of teachers' assessment awareness, and they are not "ready-made solutions or answers to problems that arise within complex and diverse classroom assessment scenarios" (Xu & Brown, 2016, p. 156.).

Therefore, I argue that TAL is best understood in practice. Involving teachers in research studies that takes them through the previously mentioned levels helps them understand the links between what they do and how they can improve their effectiveness and plan for new changes. Teachers need to participate in assessment-related activities within real communities (Adie, 2013; Lukin et al., 2004). These learning opportunities will enhance their TALiP. Teachers need to continually develop and negotiate their role as assessors (Cowie et al., 2014).

Accordingly, I intend to adopt the TALiP model as a theoretical framework for the current study because it aligns with my beliefs about TAL. The following section will explore the assessment knowledge base suggested by the TALiP model, upon which the other components of the model is built, and which will guide the research through the components of the designed research tools, which in turn will be used to explore the six components of the TALiP framework.

## **2.4 The Knowledge Base of Language Teacher Assessment**

### **Literacy**

This part of the chapter focuses on exploring the knowledge base of LTAL as per the adopted TALiP model. This section aims to come up with what constitutes each of the seven components of the assessment knowledge base suggested by the TALiP model. The outcome of this section will guide the content of the research tools, which will be designed in the current study to explore TAL in the described context, primarily the tool that will be used to explore the participants' confidence with the assessment knowledge base, which is equivalent to the first level of the TALiP model. It will also guide the other research

tools, which will explore the participants' conception of the assessment knowledge base, how assessment knowledge components are practised in their context, the participants' beliefs of and satisfaction with the APs in their teaching context, and finally how these APs can be better practised in their teaching context as per their suggested ideas.

This section will explore the following seven components of assessment knowledge base suggested by the TALiP model: (1) Disciplinary Knowledge & Pedagogical Content Knowledge, (2) Knowledge of Assessment Purposes, Content, Design, (3) Knowledge of Assessment Grading, (4) Knowledge of Assessment Feedback, (5) Knowledge of Peers- and Self-Assessment, (6) Knowledge of Assessment Interpretation and Communication, (7) Knowledge of Assessment Ethics. Brookhart (2011) argued that for each component of the assessment knowledge base, teachers need theoretical declarative knowledge and practical procedural knowledge in the form of skills. She explained that teachers need the theoretical knowledge to select, critique, or evaluate assessments, and they need the procedural knowledge to design their assessments.

#### **2.4.1 Disciplinary Knowledge & Pedagogical Content Knowledge**

Brookhart (2011) emphasised that all APs should start with knowledge of what to assess (disciplinary knowledge). Knowledge of pedagogical content knowledge was also one main component of assessment knowledge since the appearance of AFT standards in 1990. Brookhart explained: "In order to be able to assess students well and to make sound decisions based on the results, teachers must understand general principles about how students learn, and they must understand deeply the content area(s) they teach" (Brookhart, 2011:6).

First, disciplinary knowledge is awareness of language characteristics, which teachers will assess "what to assess" (Brindley, 2001; Davies, 2008; Inbar-Lourie, 2008a). For teachers to be familiar with this knowledge, they need to know about communicative competence (McNamara, 1996). Communicative language competence was defined in terms of four kinds of knowledge. 1. Knowledge of grammatical and lexical resources of the language. 2. Knowledge of feasibility of using the grammatical structures. 3. Knowledge of how to use them in social



interaction. 4. Knowledge of judgement if utterances are possible, feasible, and appropriate (Fulcher, 2010). Fulcher explained that different language models were used as the basis for test design for communicative purposes (Bachman, 1990; Bachman & Palmer, 1996; Canale, 1983; Canale & Swain's, 1980). In the 1990s, there was a reaction against these models because communicative language testing was more concerned with tasks, which were mostly performance-based (Shohamy, 1996); however, this performance-communicative task-oriented approach was criticised for not following specific, well-defined structures. As a reaction to this criticism, some performance models appeared; these model contained no reference to competences; it only included what learners "can do" in a variety of communicative situations, using functional language. The most widely spread one is the Common European Framework of Reference (CEFR) (Council of Europe, 2001).

Second, pedagogical content knowledge deals with how students learn a language and methods of facilitating their learning. Abell & Siegel (2011) stated that this knowledge is necessary for teachers performing assessment tasks because adopted language learning theories dominate language classroom pedagogy and consequently affect assessments practices. For example, according to Nassaji & Fotos (2004), when Skinner's behaviourism dominated, where language learning entailed a process of exposure, practice, reinforcement, and habit formation, the audio-lingual method was adopted, which relied on teaching grammar, vocabulary and drilling exercises. This situation led to testing discrete points of language that ignored learners' communicative competence. The dominance of the cognitive theory of language learning (Chomsky's innatism) was accompanied by the widespread of the communicative teaching approach. This adopted approach resulted in abandoning the use of formal grammar instructions or error correction and depending mainly on natural exposure to the language. Accordingly, assessment depended on communicative tasks, but learners' errors were not addressed, which in turn led to error fossilisation and learners did not reach the intended language proficiency to use the language fluently in situations outside the classroom. The appearance of Piaget's constructivism, which believed that learners are actively involved in constructing personal meaning out of their experience, promoted the notion of giving feedback on learners' performance to enable them to notice their errors and learn from

them (Ellis, 2006; Spade & Lightbown, 2008). This theory reinforced concepts, such as formative assessment and self-/peer-assessments. Vygotsky's theory, which viewed knowledge as a socially constructed domain (Lantolf, 2000) required a teacher who is capable of promoting students learning through mediation. Accordingly, teachers focused on scaffolding learning through continuous dynamic assessments that take learners from one stage to the other through mediation (Cheng, Rogers, & Wang, 2008).

These few examples of the interrelated bond between language theories, methodology, and assessment, and how they affect one another makes teachers' knowledge of language learning theories and methodologies essential, as they will interrelate with and affect their assessments' decisions and practices (Ball, Thames, & Phelps, 2008). Giraldo (2018) added awareness of learning strategies and styles as two sub-components of pedagogical content Knowledge.

#### **2.4.2 Knowledge of Assessment Purposes, Content, and Methods**

Brookhart (2011) emphasised that teachers should understand the purposes and uses of assessment and be skilled in using them. Xu & Brown (2016) believed that teachers need to know why they assess, how to relate assessment methods to learning goals and content, and relevant assessment strategies (AFT, NCME, & NEA, 1990; Brookhart, 2011; JCSEE, 2015). Xu & Brown (2016) explained that the knowledge of assessment purposes, method and content has a theoretical component and a practical one; Brookhart (2011) emphasised that assessment knowledge involves both theoretical and practical knowledge. The following part will present the theoretical and practical components of this knowledge.

##### **2.4.2.1 Theoretical Knowledge of Assessment Purpose, Content, and Methods**

###### ***Knowledge of Assessment History, Purpose/Philosophy, and Impact***

Fulcher (2010), in his practical book, addressed to language teachers/assessors, argued that some knowledge of assessment/testing history could be beneficial for teachers' APs. He added that exposure to assessment/testing history would make teachers familiar with various

assessment philosophies and their tangible impact on assessment practices adopted by different contexts, especially that many of these philosophies control learners' opportunities for further education. Shohamy et al. (2017) added that knowledge of these philosophies and impacts would empower language teachers and make them able to defend their assessment beliefs and practices against assessment policies and regimes imposed on them. Knowledge of assessment purposes could guide teachers on what works and does not work in their teaching contexts.

Out of the various assessment philosophies that dominated testing history, two philosophies stood out: assessment of learning and assessment for learning (Latif, 2017). The former views assessment and instruction as two independent activities. This philosophy gives weight to measuring and reporting learners' achievement; therefore, assessment is employed to assess representatives of learners' knowledge (William, 2011) and the outcome of what teachers teach (Elmahjoub, 2017). The focus of assessment is on the product; assessments are not providing learners with any learning experience (Richards & Renandya, 2002). Thus, the purpose is gathering information about learners' achievement for judgment (Efeotor, 2017). The data derived from such tests are meant to rank, evaluate, monitor, and place students in specific categories (Lee, 2011; Shohamy, 2001). This philosophy created 'a testing culture' (Inbar-Lourie, 2008b; Shepard, 2000), which reinforced the use of tests as a mono-assessment tool. This view of assessment is rooted in the positivistic epistemology. Positivism believes in constructing predetermined knowledge of the language that learners are required to acquire and be tested in.

The latter philosophy views assessment and instruction as an integral part of each other; therefore, assessment practices are used as instruments for facilitating learning. This philosophy focusses on what learners can produce with their acquired knowledge and not just their ability to recall and reproduce what they learned. It also focuses on providing feedback on learners' performance and consequently enhancing learning by revealing points of strength and weakness for both students and teachers in order to adjust both learning and teaching (Brookhart, 2009; Ecclestone, 2012; Mohan, 2011). Students' evaluation is provided in the form of a profile rather than a numerical score (Birenbaum, 1996). This philosophy promoted "an assessment culture". Contexts, which adopts

assessment culture, view learners as active, empowered partners in the assessment process, who monitor their learning, provide feedback to their peers and set criteria for evaluating progress. Teachers' role within this culture is scaffolding learning based on feedback obtained from different varied assessment tools. Assessment culture emphasises formative assessment practices, (Black & William, 1998), and introduces suitable assessment tools such as Dynamic Assessment (DA) (Lantolf & Poehner, 2008). This view is grounded in the interpretive epistemology, which views knowledge as a socially constructed reality that is dynamic constantly reshaped by teachers and empowered students (Looney, 2008; Shute, 2008).

### ***Knowledge of Assessment Methods***

Brookhart (2011) emphasised that teachers must know what kind of assessment method they can use to draw out the knowledge and skills they intend to measure. Stiggins (2008) calls this "target-method match". Stiggins (2008) categorised assessment methods into four kinds: selected response, essay, performance assessment, and personal communication. Teachers should understand how to write or select appropriate high-quality assessment tools to match any of these methods.

Richard & Schmidt (2010) & Atac (2012) classified assessment tools into formal/traditional and informal/alternative. Popham (2003) & Reeves (2007) associated formal and informal tools to the purposes they intend to achieve; they emphasised that formal assessment tools are mainly conducted for summative reasons to evaluate students, while informal assessments are mostly conducted for formative purposes to improve learning. Formal/traditional tools refer to predetermined testing measures such as selected-response tests (multiple-choice questions, true/false questions, matching questions), brief constructed-response tests (short-answer questions), and essay questions (Brookhart, 2011). Formal tools take the form of graded tests or exams, such as (1) placement tests used to place learners into classes or levels, (2) achievement tests used to discover how much learners achieved, (3) diagnostic test used to diagnose difficulties that learners may have (Brown, 2005). Teachers can also use other assessment tools to motivate learners to study through quizzes or short tests.

On the other hand, informal tools can take the form of conferences, debates, demonstrations, diaries, journals, dramatisations, exhibitions, games, observation, peer assessments, projects, story retelling, portfolios, self-assessment, and think aloud. These tools involve authentic assessment tasks (Airasian, 2005; McMillan, 2007). Their results can be recorded in anecdotal records, checklist, learner profiles, progress cards, questionnaires, rating scales (Tsagari, 2004). Bachman (2005) recommended the use of alternative tools, as they are informative; they can be gathered over some time, provide positive feedback, and finally improve instructions. Coombe et al. (2012) emphasised the necessity of using multi-measure assessments because using one single tool cannot provide the assessors with all necessary information on learners' language performance or accurately measure their ability and proficiency levels.

### ***Knowledge of Assessment Content***

Brindley (2001) & Davies (2008) involved knowledge of assessment content/development/design as the main component of assessment literacy. Knowledge of assessment content is equivalent to Inbar-Lourie's (2008a) "how to assess". This knowledge involves knowledge of assessment qualities, types, development stages, and cognitive taxonomies required for item design.

### ***Knowledge of Assessment Qualities***

Bachman & Palmer (1996) and Weir (2005) considered an assessment tool to be useful if it has six qualities: construct validity (assesses intended specification); authenticity (assesses representation of real performances); interactiveness (involves test-takers demonstrating strategic competence); impact (has beneficial washback); reliability (uses consistently clear prompts and scoring instruments); and practicality (uses available resources that meet the assessment's needs).

## ***Knowledge of Assessment Types***

There are different types of assessments, which are categorised based on different criteria (McMillan, 2000). The flowing table summarises assessment types:

Criteria	Assessment types	
<i>According to Assessment Source</i>	Externally-Mandated	Internally-Mandated
<i>According to Assessment Purpose</i>	Summative	Formative
<i>According to the Results Use</i>	Norm-referenced	Criterion-Referenced
<i>According to Assessment context</i>	Large-Scale Assessments	Classroom-Assessments

*Table 2.4 Types of Assessments*

*Internally-/Externally-Mandated:* Internally-mandated assessments are set by teachers or context administration (Davidson & Lynch, 2002). They are "ecologically sensitive"; their outcome, whether in the form of scores or performance, are interpreted to a specific learning environment (Cronbach, 1984). The externally-mandated assessment comes from outside the local context (Fulcher, 2010). They are summative. Data about learners do not feedback into teaching or learning. They are of high stake nature, as their results can have a severe impact on learners' plans, management, and teachers. Unfavourable results can lead to reforms.

*Summative/Formative:* summative assessments are used to evaluate learning at the end of an instructional unit by comparing their progress against some standards or benchmarks (Stiggins, 2002; Derrich & Ecclestone, 2006). On the other hand, formative assessments are used to monitor students' learning and provide ongoing feedback that can be used to improve teaching and learning (Threlfall, 2005).

*Norm-/Criterion-Referenced:* Norm-referenced are used to discriminate between test-takers; scores obtained from such tests reflect the position of an individual to others (Fulcher, 2010). Tests are seen as measuring tools of knowledge, skills, and abilities that are stable parts of the test takers. Numerical

values evidence the existence of such knowledge. Criterion-referenced is used to make decisions about whether a test taker has achieved a pre-specified criterion or standard that is required for a particular context (Stiggins, 2001). The results such assessment are expected to aid in making decisions about learning and instructions.

*Large-Scale/Classroom:* In large-scale assessments, tests are used on a national or international scale. Designed tests should be appropriate for their stated purpose and target population. Large-scale tests require an intensive investigation on validity and reliability matters to determine the extent to which score meaning can be generalised (Fulcher & Davidson, 2007). The context of a large-scale test is the testing environment. Physical factors might have an impact on the candidates' performance. Classroom assessments refer to "those formal and informal procedures that teachers employ to make inferences about what their learners know and can do" (Popham, 2009: p.6). Teachers can tell a great deal about learners' knowledge, abilities, and skills. In classroom assessments, the context is the learning environment, which is constructed of a set of learning experiences that are designed to lead to language acquisition and communication; the context is part of the construct; it is not construct-irrelevant. Classroom tasks take place over some time (Moss, 2003).

### ***Knowledge of Development Stages***

The assessment development process requires integrated efforts as it goes through different stages (O' Sullivan, 2011): (1) Planning and designing stage, which involves setting the purpose, construct, target population, format, and response format (2) Developing stage, which involves writing assessment specification: title of the specification, purpose, general description, detailed description of what the test taker encounters, detailed description of how the test taker will provide the answer, the criteria for evaluating the response, sample or examples for the task, and task-scoring process (Davidson & Lynch, 2002). (3) Administrative and monitoring stage, which involves considering physical conditions: setting and monitoring test structure and timing, securing test delivery, storage, invigilation, checking the availability of human and material resources, and planning for sudden changes to improve test practicality.

### ***Knowledge of Cognitive Taxonomies***

Brookhart (2011) claims that teachers should be aware of learners' thinking skills and be familiar with different taxonomies (e.g., Bloom, Webb), in addition to understanding kinds of cognitions these taxonomies describe. She explains that knowledge about learners' thinking skills will enable teachers to design different test items and performance tasks that meet them. Brookhart added that teachers need this knowledge even when formulating simple classroom questions or evaluating curriculum, material, test items, and tasks written by others.

#### **2.4.2.2 Practical Knowledge of Assessment Purposes, Content, and Methods**

Teachers need specific skills to design assessments; they need to acquire specific skills such as setting purpose, writing goals, aligning goals with instruction, defining constructs, writing specification, writing items, enhancing layout, writing test syllabuses, incorporating technologies. (AFT, NCME, & NEA, 1990; Brookhart, 2011; JCSEE, 2015). Design cycle for tests differs from task-based tasks (Fulcher, 2010). The following illustrates the skills required for designing formal tests.

##### ***Test Purpose***

A statement of test purpose should include information on the target population, their level, the domain of language use, language abilities, in addition to a justification of constructs and content selection (Fulcher & Davidson, 2007)

##### ***Construct Definition***

English language assessments assess what is called language proficiency or communicative language ability (Celce-Murcia et al., 1995; McNamara, 1996). These terms are abstract ones that need to be changed into constructs to measure them. Constructs are usually described in models. Models are used as a theoretical description of what it means to be able to communicate in a second language. Test developers need to link the constructs to the purpose and context of the test; which is part of test framework; frameworks are used to refer to the selection of skills and abilities from a model that is relevant to a specific assessment context (Chalhoub-Deville, 1997). This process means that during



assessment design, test developers work on three levels. On the first level, test developers work on selecting a particular theoretical model; then they go to the second level, where they choose their assessment frameworks, which will lead to the third level, on which they build their assessment specifications.

### ***Test Specifications***

Test specifications are generative explanatory documents for the creation of test tasks (Fulcher & Davidson, 2007). They also tell the rationale behind all choices made. Another purpose of a test specification is to create "test forms" that originate from the same specification to secure equivalency, reliability and validity (Davidson & Lynch, 2002). Mislevy, Almond, & Lukas (2003) listed all possible specifications that test developers might need for a test: (1) Item/task specifications, this describes the prompts that are designed to elicit the evidence. Item specifications state required material, the feature of the prompts, instructions, and sample items. (2) Evidence specification, this is a description of what a test taker is expected to do "response attribute", and how the response will be scored. (3) Test assembly, this provides instructions on how the entire test is constructed, such as number of item types, number of items required for each category, and the minimum number of items needed to meet the target reliability. (4) Presentation specifications, this tells how items and support material are presented to test taker. The presentation involves the margin size, the font type and size, spacing, where page numbers will appear. (5) Delivery specification, this specifies the set of details for test administration. Test administration includes test security, timing, spacing, number of invigilators, allowed objects, and allocated time.

### ***Item Writing***

Writing test items should be built on a test specification. Choosing the item method is crucial because this choice can affect the candidates' scores because some item methods are appropriate for testing some abilities but not others (Allan, 1992). It is recommended to use more than one item method to test the same ability or objective. Item writers need to be aware of features, advantages, and disadvantages of test items to avoid negative washback of the test (Heaton, 1998). Alderson et al. (1996) presented objective and subjective item types that could be used in language tests. Objective /closed-ended items can include but

not limited to: multiple-choice, dichotomous items, matching, information transfer, organising tasks, editing, gap-filing, cloze, C-test, dictation, and short answer questions. Subjective/open-ended items can take the form of composition/essays, summaries, or oral interviews. Hughes (2003) suggested certain practices that can enhance item writing and accordingly enhance assessment reliability, such as constructing well laid out items, writing clear instructions, providing examples, including many items, and writing test syllabus to inform test users of test formats.

### ***Prototyping and Field Tests***

Prototyping is the process of evaluating test items/tasks; sometimes, this process is referred to as "moderation" (Fulcher & Davidson, 2007). It takes place before test forms are designed. Test designers are expected to respond to any concerns that might show up. Field testing could test samples on users from the target audience.

### ***Knowledge of Assessment Administration***

Different administration procedures follow tests design; however, they differ from one context to the other depending on the level of test formality (Fulcher & Davidson, 2007). These steps focus on (1) test forms preparation, (2) sourcing and collecting raw materials, this includes reading passages, recorded listening texts, graphs, illustration, and charts for information transfer items, (3) item review, this process requires a committee to check items against test specifications, (4) item banking, all items that survived review need to be stored in a format that allows easy access, (5) test assembly, items are extracted and combined to make a linear form, (6) preparation, the layout should be uniform across forms presented in the same font type, style, and size, (7) printing and duplicating, this involves making copies of test papers, answer papers, or multiple copies of tapes, (8) storage systems; in this stage, security is critical, (9) record keeping, the test forms need to be referenced according to the intended place and time, (10) distribution systems, paper-based tests need to be extracted from storage and dispatched by secure means to test sites, (11) delivery systems, tests need to be transferred safely to test rooms. All invigilators need to be trained to deal with unforeseen events such as late arrivals, disruptive test-takers, potential cheaters, those who finish early, and those who need to use the

restroom. They also need to be trained on how to check the identity and give instructions.

A different design cycle would be followed when designing classroom-based assessments or tasks (Fulcher, 2010). If teachers need to design alternative assessments, they need to decide on:

### ***Assessment Purpose***

The first step in designing a task-based assessment is setting the purpose. The purpose could be to assess learners' achievement in class on material covered in a specific time, provide a summative evaluation on learners' achievement of a course, diagnose learners' weakness, or give feedback on learners' progress. The task could be simply designed to assist and scaffold learning by creating opportunities for learners to discover and assess their own or their peers' strength and weakness (Popham, 2009).

### ***Assessment Criterion***

There should be an assessment criterion related to the course objective to address the assessment purpose. Teachers can sample content directly from the syllabus or design their tasks. Teachers could design tasks similar to real-life ones, such as reading manuals, following instructions, or writing reports depending on the context and purpose of the assessment. In this case, to describe the assessment criterion, the assessment developer should collect representative samples of the target skills, knowledge, behaviour, and abilities, or interview people involved in the target tasks to discover what kind of target behaviour is expected from the test taker to come up with. Without this explicitness, we would have design chaos (Fulcher, 2010).

### ***Test Specifications***

Teachers are required to clearly describe the knowledge and skills to be tested in the test framework. To write valid test items, teachers need to write 'Item-objective congruence', which refers to the relation between the item/task and the learning objective that it is designed to test. (Davidson & Lynch, 2002). Davidson & Lynch designed a simple classroom test specification. It starts with a general description that links the task type with the syllabus objective. The prompt

attribute defines what instructions the learners will be given and what kind of input is required to generate the required response. It could also contain information related to the source and difficulty of input materials. The text type, ranges, and genres may be specified to link them directly to a criterion context. The response attribute describes precisely what the learners are expected to do in their response to the prompt, such as specifying an extended piece of writing or production of an extended speech. All specifications should contain sample items, sometimes "anti-items" to show what is not intended. Including specification supplement is also helpful, as it includes details of any additional information needed to construct items such as specifying precisely the grammar forms to be tested, adding a vocabulary list, or adding textbook from which reading texts were selected.

Despite the beneficial washback of using test specification in designing classroom assessments, many teachers did not like these detailed specifications. They find them lengthy and time-consuming; others believe that they limit creativity and instructions rather than enriching it (Popham, 1994). In reaction to these arguments, Popham suggested that for pedagogical purposes, specifications could be "boiled down", just a general description and a sample item.

Another challenge that classroom teachers would meet is designing the specification for authentic tasks if authenticity means "replicating real life" in performance assessments. Bachman & Palmer (1996) suggested describing the task according to features that exist in the target language use situation across several categories: the facets of the testing environment (place, equipment, participants), the facets of the test rubric (organisation, time, instructions), the facets of the input (format and language), the facets of the expected response (format, language, restrictions on the response), and the relationship between input and response (whether reciprocal, non-reciprocal, or adaptive).

Knowledge and skills involved in assessment design were thoroughly discussed because it was emphasised that language teachers need to pay thorough attention to the assessment design process while designing assessment to produce valid, reliable, fair assessments (Brookhart, 2011).

Cohen (1994) suggested using performance-based tasks, where learners are given opportunities to express opinions as well as demonstrate both communication and comprehension abilities. These tasks make learners use what is learned, and they are transferred from controlled learning to real-life performance. (Hughes, 2003; Weir, 2005) warned against the use of fixed response format in the form of multiple-choice questions and true/false questions, despite being widely used in reading/listening assessments because they test receptive knowledge, ignore productive skills, train guessing rather than learning, and might include more than one correct answer. Learners should be required to accomplish tasks that reflect their comprehension and give justifications for their answers and interpretations. Different answers should be accepted since individuals' interpretations of the same input are different. Shohamy (1988) suggested using a combination of assessment methods for assessing speaking to obtain an overall picture of oral proficiency. These methods can comprise personal conversation, detailed description, topical discourse, situations, or giving directions. Depending only on interviews, despite their high face validity as elicitation tools for communicative language assessment, is limiting due to the social distance between the interviewer and the interviewee (Perrett, 1990). Assessing writing skill should take into consideration the writing process and not only the writing product (Cohen, 1994) since the finished product needs to emerge after a series of drafts, in which the written piece takes shape after successive evaluations and revisions based on provided formative feedback.

### **2.4.3 Knowledge of Assessment Grading**

Brookhart (2011) stated that teachers' knowledge about constructing scoring schemes that quantify learners' performance is unavoidable. This knowledge is essential for them to be able to make useful inferences, which would lead to improved learning.

For classroom assessments, teachers should know and be able to use different grading methods for items or tasks (answer keys, checklists, rubrics, rating scales). They also need to be confident with using aggregating scores into meaningful composites in the form of points, percentage, grades, or proficiency levels. "Their understanding should include basics of simple linear scaling, weighting components, and precision of the results" (Brookhart, 2011. P.9). Lack

of grading knowledge leads to invalid assessment and grading practices (Brookhart, 1999). Therefore, grading knowledge is considered fundamental for teachers performing assessment tasks (Campbell & Collins, 2007). The emphasis is on teachers' ability to construct scoring schemes, not just performing grading or scoring tasks.

In addition to being competent at designing closed-response grading tools mentioned ahead, teachers should be able to construct rubrics or rating schemes for open-response items such as writing or speaking performances (Fulcher & Davidson, 2007). Hamp-Lyons (1991), introduced different types of scoring: (1) Holistic scoring: depends on assigning a single score to a script based on the overall impression of the rater. It reflects the authentic personal reaction of a reader or a listener (Weigle, 2002); however, it does not take into consideration inner components or sub-abilities, and the scores are not easily interpreted because it lacks detailed feedback (Mckay, 2006). (2) Analytic scoring: it is a reliable diagnostic technique because the script or the speech is rated against several aspects/subskills; but it is time-consuming and requires trained raters (Weir, 1990). (3) Primary Trait Scoring: scoring is defined for a specific ability. It is essential to precise the abilities and sub-abilities within this narrow trait (Fulcher, 2010).

To defend their grading and scoring practices, teachers can use several statistics to help justify their tests' scores, especially for closed-ended responses. They can test item difficulty (item facility) by calculating the proportion of test-takers who answer an item correctly. Another way of defending the reliability of their grades is by testing the ability of the item for discrimination. This concept is based on the assumption that the responses to individual items are capable of discrimination between higher ability and lower ability of the test taker. For open-ended responses, Hughes (2003) suggested double-rating and calculating agreement between raters; he also recommended training raters and having agreement meeting before grading practices.

For task-based assessments, where teachers are less interested in scores, and they are mainly interested in judging whether the test taker reached the criterion or the objective, scores and scorer reliability is not a problematic

area. Teachers are interested in the "decision consistency of the assessment"; this refers to whether the assessment is consistent in deciding if the learner has reached the criterion. The purpose of such an assessment is to provide detailed explicit feedback to learners; in this case, no numeric score is required. Rea-Dickens (2006:168) described the quality of feedback teachers should give to their learners. It should be descriptive rather than evaluative, it should highlight aspects that need improvement and means of improvement, and it should provide learners with enough time to digest and respond to teachers' feedback. For these tasks, feedback and grading are not separable from one another. Feedback involves responding to learners' performance in light of pre-set criteria with the purpose of learning and improvement (Hattie & Timperley, 2007; Rea-Dickens, 2004).

For external assessments, especially standardised ones, to which teachers are required to respond through grading, teachers should have the knowledge and skill that enable them to interpret scores obtained from them (Brookhart, 2011). Other practitioners believe that teachers do need to know how to design grading schemes and calculate scores validity for such tests; they need to know about standardised assessments as much as they know about classroom assessments grading because such standardised tests influence both teaching and learning (Guskey, 2007; Leighton, Gokiert, Cor, & Heffernan, 2010; McMillan, 2003; Shepard, 2006). Teachers should be able to use these understandings about score meaning to improve students' learning. Brookhart argued: "This understanding should support teachers' skills in navigating some of the pressures that come with current accountability policies" (Brookhart, 2011: p.10). Teachers prepare students for these tests, so it is also apparent to know their grading techniques.

Teachers' role in grading external assessment should not be limited to an administrative role in the form of scoring (Campbell & Collins, 2007). Teachers should be able to interpret norm- and criterion-referenced scores in the form of:

- Understanding measurement error and confidence intervals;
- limiting generalisation to the construct assessed and not beyond;
- understanding the difference between grade-equivalent scores and

grade-level instructional objectives; and understanding longitudinal and cross-sectional analysis. (Brookhart, 2011:p.10)

#### **2.4.4 Knowledge of Providing Feedback**

Brookhart (2011) argued that teachers should have the knowledge and skills that enable them to provide useful constructive feedback on learners' work. Effective feedback involves providing learners with information about their work against criteria that were shared with them as part of their learning objectives (Hattie & Timperley, 2007). Feedback should be elaborated, specific, and different from one learner to the other (Shute, 2008). Learners should be able to identify required areas of improvement and work on them without having the comment too explicit to the extent that they feel that work is done for them (Kluger & DeNisi, 1996). Writing clear and useful feedback requires much practice and training (Lee, 2009).

Feedback is associated with two pedagogical approaches: Assessment for Learning (AFL) and Dynamic Assessment (DA) (Fulcher 2010). AFL starts with teachers setting goals, constructing tasks that will move learners towards these goals, and finally evaluating how well the learners have succeeded in achieving them (Cumming, 2009). In AFL, results are given in the form of feedback that is meant to modify the learning process, not a score (Leung, 2004; Leung & Scott, 2009). In DA, teachers describe the gap between the learners' current stage of development and the next stage of development, without necessarily showing the final target of performance (Shepard, 2007). In DA teachers are mediators. They intervene in the learning process to modify learners' use of language. There are two types of mediations, "interventionist" and "interactionist". The former standardises the mediation, so it is shared across learners; in the latter, the mediator interacts with each learner depending on their current stage of development (Lantolf, 2009:363). Fulcher (2010) explained that there are three methods closely related to DA, "graduated prompt", "testing the limits", and "mediated learning experience". In the "graduated prompt", the mediator creates a task with a graded series of questions for learners who have a problem completing a task; the questions focus on the problem. In "testing the limits", a learner is given feedback on their performance on a task and then asked to verbalise the problem they have and what he/she will do to overcome it. In the



"mediated learning experience", the mediator interacts with the learner to make him/her move toward the next stage of learning through scaffolding attempts to communicate.

#### **2.4.5 Knowledge of Self-and Peer-Assessments**

Brookhart (2011) emphasised that teachers should be trained on involving learners in their assessments by assisting them in using assessment feedback to enhance their assessments and their peers' assessment to be able to make sound educational decisions. She stated that teachers should be familiar with self-and peers-assessment principles that focus on their roles in assisting learners to interpret assessment results, track their learning, communicate about their learning, and plan next steps in their learning. Teachers should be able to motivate learners to have control over their learning through self-regulation and benefiting from their assessment experience. This case is different from formative feedback provided by teachers because as per self-and peer-assessment principals, teachers will direct learners to use assessment information. If teachers failed in assisting learners in doing so, then the formative assessment is not sufficient (Furtak et al., 2008; Torrance & Pryor, 1998). When teachers explain to learners the principles behind adopting self-or peer -assessments, the adopted practice will support learning and also support students' feelings of self-efficacy and control (Bandura, 1997; Ryan & Deci, 2000) and self-regulation (Butler & Winne, 1995). Understanding these connections helps teachers select and use formative assessment strategies and work with students appropriately (Moss & Brookhart, 2009).

Black et al. (2002, 2004) suggested using self- and peer-assessments to promote learners' ability to notice the gap in their learning process; they recommended providing learners with the criteria teachers use in judging the quality of their work; moreover, learners could be asked to produce their rating criteria in groups and use them to peer-assess one another. Different practical practices are suggested to make learners keep records of their work and how their work improves and develops. They can keep a diary, a continuous card, a digital audio, a video diary, an online blog, in which samples of work and commentary are saved. This practice naturally leads to the use of portfolios, where learners are expected to collect samples of writing or digital copies of their

speech; they may also contain listening and reading texts with an assessment of how well they were understood, and reactions to them; these portfolios could be assessed by teachers, students and peers (Fulcher, 2010).

#### **2.4.6 Knowledge of Assessment Interpretation and Communication**

Teachers need to know ways of interpreting evidence derived from the norm- and criterion-referenced assessments, in addition to ensuring that their interpretations have sufficient evidence (Brookhart, 2011). Teachers should be able to understand "measurement error and confidence intervals, limit generalisation to the construct assessed and not beyond, understand the difference between grade-equivalent scores and grade-level instructional objectives, understand differences between scores for individual students and class- or school-level aggregated scores, and finally understand longitudinal and cross-sectional analysis" (Brookhart, 2011:10). Teachers should be able to use these understandings about score meaning to improve students' learning.

Popham (2011) added that teachers need to understand at least basic concepts of reliability and validity to be able to evaluate their assessments or externally- mandated ones, to which they are exposed. Teachers need to investigate different types of assessment validity. Construct validity could be checked through examining the relationship between assessed constructs and assessment purpose, context, and population. To judge content validity, teachers can examine test specifications. On the scoring level, they need to make sure that scores are given only to responses related to the assessed skill/sub-skill. Teachers can also check criterion-oriented validity by comparing assessment results to results provided by another independent and highly dependable assessment measuring the same ability or skill (Hughes, 2003).

To investigate reliability-related matters, Brookhart (2011) mentioned that teachers need to know how to interpret data related to test design, such as item difficulty and item discrimination; they should also be able to interpret data from large-scale tests, namely means, modes, medians, and bell curves. They should have the ability to calculate them and infer students' strength and weakness to communicate inferences to different stakeholders.

Teachers should be able to articulate their interpretations of assessment results and their justification regarding assessment decisions to different stakeholders: learners, school, and community (Brookhart, 2011; Stiggins, 2008). To perform these assessment tasks, they need to be skilled in communicating results and means of improvement to other stakeholders in the form of report cards or any other method that can justify their decisions about learners' performance and intended follow up. They should be able to participate in counselling committees, at which they assist learners with guidance regarding other educational decisions. Finally, teachers should be able to participate in school committees about assessment-related issues and their impact on reforms related to assessment design, curriculum, materials, grading, or reporting (Stiggins, 2008).

#### **2.4.7 Knowledge of Assessment Ethics**

Teachers need to understand the legal and ethical responsibilities associated with their assessments and their impact on learners (Brookhart, 2011). Teachers are required to know how to work towards equity, non-discrimination, inclusion, and social justice. They need to be exposed to principles of critical language testing that encourage giving voice to teachers and students about their assessment (Tierney, 2013). This knowledge minimises the negative consequences of assessments (Stoynoff, 2008). It is the learners' right to be treated ethically and fairly and not to be harmed (Lynch & Shaw, 2005). Ethicality includes the issues of harm, consent, fairness, deception, privacy and confidentiality, validity, absence of bias, access, administration and social consequences (Kunnan, 2003, 2004). A call for test ethicality is not a call against tests (Shohamy, 2001); it is about the role that tests play in societies and its misuse. Assessment ethicality is a call for quality assessments that require shared authority, collaboration, involvement of different stakeholders, as well as meeting various criteria of validity.

Traditionally, a fair test was viewed as one that is free from bias, discrimination, and favouritism (Tierney, 2010). Evidence of fairness was addressed via statistical procedures such as validity and reliability (Camilli, 2006; Volante, 2006). Another conceptualisation for fairness that went beyond statistics included defining a clear purpose for assessments, developing specifications,

evaluating content and conducting a field test examination (Plake & Jones, 2002). However, this approach did not tackle all ethical needs due to the diversity of assessment purposes, types, population, and procedures. This situation led to the appearance of "professional morality" (Davies, 2004). Professional morality takes the form of a code of ethics and practice, to which the community of assessment practitioners yield.

Practically, teachers should avoid construct-irrelevant, controversial, inflammatory, offensive, or upsetting test material (Pitoniak et al., 2009). They need to provide equal chances to learners, especially in classroom assessment. Teachers are required to justify their decisions, negotiate concepts, such as validity, reliability, and washback (Jia, 2009), and use multiple methods (Troudi et al., 2009). These practices would allow learners to show their knowledge and reveal a composite picture of their learning (Earl, 2003). Finally, since tests are no longer viewed as innocent tools but rather instruments that play central roles for people, education, and societies (Shohamy et al, 2017), it is the assessors' responsibility to create assessments that are more inclusive, democratic, just, open, fair, equal, and less biased. Shohamy et al. (2017) explained that teachers need to design assessments that are valid and reliable, where validity includes the protection of the personal rights of others. Part of assessment ethicality is to ensure its role as a supportive tool for learning and teaching to maintain an acceptable level of positive washback that is considered one key feature of valid assessments.

In Section 2.4, I tried to present details on assessment knowledge base; however, the assessment knowledge base cannot be dealt with as prescriptions or ready-made recipes that could be taken to different contexts. Teachers teach English for different purposes in diverse contexts, which are governed by diverse, varied assessment philosophies, beliefs, and practices. Assessment theoretical and practical knowledge cannot be applied directly in different contexts. There should be a bridge between such knowledge that is informed by research and theoretical epistemologies and APs. Teachers as assessors need to act as bridges that allow only appropriate knowledge/practices to access learning contexts to ensure fair, ethical effective assessment practices. Teachers' conceptions, beliefs, and affective dimensions allow them to engage in

interpretive dialogues through reflective practice to filter what works and what does not work in their social domain. Therefore, the current study will explore teachers' confidence with this knowledge base and also their need for such knowledge. It will also explore teachers' beliefs about how this assessment knowledge base is and should be practised in their contents. I believe the finding of the study will provide an understanding of TALiP, in addition to opportunities for development in such knowledge as a result of professional dialogues that will take place with the teachers/assessors participating in the study to assist them to reach an awareness of their current assessment identity. This professional involvement will enable them to see their points of strength and weakness and consequently reflect on future development they need in the assessment field.

The next section will explore similar empirical studies that were conducted with the same aim. These explored studies could provide rich information on research methodology, tools, and findings, from which the current study can benefit. The findings of such empirical studies would be referred to during the findings and discussion phases because they can validate or contradict the current research findings, which in both cases will be an addition to literature related to TAL.

## **2.5 Empirical Studies Conducted on Exploring Teachers' Assessment Literacy**

Different empirical studies were conducted in different countries to explore LTAL. Although they all share the same aim, they had different research approaches and used different research tools.

In Europe, a study was conducted by Vogt & Tsagari (2014) in seven European countries to explore AL of foreign language teachers and identify their needs through both questionnaire and interviews; it was found that not all elements of LTAL were developed. Vogt & Tsagari argued that teachers seem to learn about assessment on the job or through teaching material; the majority received either a little or no training before performing their assessment tasks. Their participants expressed a need to receive training on all assessment features. In the same European context, Fulcher (2012) explored language teachers' assessment needs using a questionnaire that was delivered over the

internet for those who had an interest in LA; he found that teachers need training in knowledge, skills, and principles of assessments to be able to respond to both classroom assessment and standardised ones.

In Canada, Huang (2018) explored teachers' training, experience, skills, and beliefs about assessment using surveys and interviews with 35 EFL teachers across 14 institutions; the study revealed lack of AL among teachers and a need to direct practical instruction in assessment with specific topics such as, assessment philosophy, design, validity, reliability, and reporting.

In South Africa, Kanjee & Mthembu (2015) explored LTAL using questionnaires, observation, and interviews; they found that teachers' understanding of summative assessments was high, while they demonstrated a poor understanding of formative assessments. The study recommended enhancing TAL through well planned professional development opportunities.

In China, Lan & Fan (2019), conducted a study to explore EFL teachers' AL and needs for training through a questionnaire sent to 344 EFL teachers, they found that TAL was at the fundamental level, and they required training and development on both theoretical and procedural levels.

Similar studies were explored in the Middle East in different EPPs, and they showed almost similar results. In the Iranian context, Firoozi, Razvipour, & Ahmadi (2019) conducted a study to explore TEL through in-depth interviews with 15 EFL teachers; they found that teachers' current conception regarding assessment is not sufficient. The participants required training in designing rubrics and formulating items. In Saudi Arabia, Latif (2017) explored EFL teachers' assessment literacy; he found that they recognise the vital role of assessment, but their training for and understanding of assessment are diverse, unclear, and inconsistent. Most instructors showed preferences towards informal assessment; however, they lacked understanding about the employment of appropriate methods. The majority were unclear about test notions like validity, reliability, test washback, and test fairness. They all believed in the significance of professional development; however, almost no steps were taken to improve professional knowledge. Latif emphasised the necessity of pre-service teachers'

training and the need for a sound academic background in assessment and testing to enable them to distinguish comprehensive and flawed assessment.

Similarly, in the Libyan context, Elmahjoub (2017) found that teachers' assessment knowledge is not derived from official preparation; they are influenced by their perception of assessment and experience with assessments on the job. In Algeria, Djoub (2017) found that teachers' views about assessment reflected a lack of AL. In the Kurdish region in Iraq, Ismael (2017) found that teachers have not been exposed to recent advances in LA; their belief and practices reveal the limited assessment literacy possessed to the Kurdish EFL teachers.

Another stream of studies focused on teachers' inclusion in assessment decisions. It was found that excluding teachers from assessments would threaten the validity and transparency of assessments and would make them feel powerless, which leads to both lower job satisfaction and self-esteem (Mansory & Meccawy, 2017). Mansory & Meccawy attributed the limited involvement of teachers in assessment to the following: illiteracy and inability of teachers to produce quality test items, enhancement of test security through limiting numbers, lack of time and financial support for teacher professional development, and interference of internal policies. Troudi et al. (2009) explained that these exclusion practices could lead to depriving teachers of their right to have their voices heard. Alnahdi (2014) supported teachers' inclusion; however, he associated teachers' involvement with proper assessment training.

## **Conclusion**

Exploring LTAL literature and different empirical studies conducted on it informed me about the necessity of conducting the current research study. First, I have not come across any study in Kuwait that explored LTAL from a socio-cultural perspective taking into consideration different conceptual, emotional, and contextual factors while exploring TAL. The only study that was informed by a similar perspective was conducted in Saudi Arabia by Latif (2020), where he used both quantitative and qualitative methods to explore EFL teachers' assessment knowledge base, beliefs, and practices.

Second, all explored empirical studies had one recommendation in common; they all necessitate teachers' development in the specific area of AL. I believe that the solution to this problem is not only external through offering teachers external ready-made practical or theoretical training. The solution to this problem needs to be tailored inside its context. It should be based on the specific requirement of its stakeholders. I believe in involving the actual stakeholders in exploring their problem, noticing it, suggesting solutions, and deciding on further steps that assist them in solving context-related challenges that could not be seen by anyone else as much as they do. Researching with this perspective does not focus only on exploring LTAL; it is a development practice, in which both the participants and the researcher develop as a result of being engaged in such professional dialogues about AL.

Third, the explored literature resulted in shaping the theoretical framework of the study, which I will adopt while investigating AL in the study, and which aligns with my research philosophy and purpose. TALiP was selected as a conceptual framework for exploring the participants' AI for two reasons. First, this framework goes beyond a focus on AL knowledge base (knowledge/skills/principles) to a consideration of dynamic affective and contextual factors that affect TAL; therefore, it provides the opportunity for a better holistic contextualised understanding of TAL. Second, the aim of the study is not only understanding TAL; it involves reconceptualisation of taken for granted concepts, in addition to providing opportunities for teacher learning and identity construction as assessors. This aim echoes the type of engagements recommended by TALiP framework, as it requires the assessors to be involved in a learning community, where they have a common language to share, negotiate, and make decisions about their APs (Lukin et al., 2004; Wyatt-Smith et al., 2010). The adopted framework assisted me in phrasing my research questions and research methodology. Besides, the explored assessment knowledge base guided me on the content of my research tools, which will be discussed thoroughly in the following chapter.

Therefore, I aim to understand, re-conceptualise, and possibly develop the AL of eight EFL teachers/ assessors working in the described context. The study seeks answers to the following research questions, which address the



components of the TALiP model. Each question takes the assessors and the researcher a step ahead towards their conceptual understanding and development of AL.

1. A. How were the assessors prepared for their assessment tasks?
1. B. How are the assessors involved in assessment tasks in their teaching context?
2. A. How confident are the assessors with the assessment knowledge base suggested by the TALiP framework?
2. B. What are the assessors' views regarding their need for the different components of assessment knowledge base suggested by the TALiP framework?
3. A. How are assessments practised in the assessors' teaching context?
3. B. What are the assessors' views about their context-adopted assessment practices?
4. How can the assessors negotiate between their views about assessments and their context-adopted assessment philosophies and practices?
5. How far have the assessors' assessment literacies been developed as a result of taking part in the research?

## **Chapter Three - Research Methodology**

### **Introduction**

To address the research purpose and questions presented in the previous chapter, I conducted an exploratory interpretive research study with eight EFL teachers/assessors in the described context. I collected qualitative data through an open-ended questionnaire, a structured interview with an open-ended checklist, a semi-structured interview with open-ended exploratory questions, a structured interview with an open-ended report, and an unstructured interview with open-ended questions. During the data collection process, the participants reported information about their theoretical and practical preparations as assessors and their current assessment roles. They also self-reported their confidence with the assessment knowledge base proposed by the TALiP framework and their needs for its different components. They described their context-adopted assessment philosophies and their views about them. Besides, they shared their critical perspectives on how they could compromise between their beliefs and their context-adopted assessment philosophies/practices. They also reflected on how the research contributed to forming their new EAP assessment identity. I subjected the collected data to interpretive analysis, which provided answers to the research questions and therefore addressed the research purpose.

The current chapter presents the research methodology; it describes the research paradigm, methodology, method, and tools. It also explains in details the research procedures, which involve how the participants were sampled and a description of the data collection procedures and piloting. Research issues are justified in the current chapter; this includes a description of ethical considerations and means through which I maintained credible, rigorous practices to assure research trustworthiness.

### **3.1 Research Paradigm**

An interpretive approach informed the current study since its principles echo the research purpose and nature of reality sought by the research questions. The interpretive philosophy is after understanding of social

phenomena and the interpretation that actors themselves give to the explored phenomena (Cohen et al., 2007; Crotty, 1998). According to an interpretive view, the research aim is to understand and explain explored phenomena as per the interpretations of the participants (Cohen et al., 2007:19) as the existence of social phenomena is not independent of their performances. Similarly, the current research aims to understand the participants' AL through involving them in a self-reporting practice, where they report their knowledge, beliefs, and views about the explored phenomenon, which can be uncovered only through their interpretations. Hence, the sought constructed reality is the product of the participants' consciousness and mind, which is a subjective nominalist view of the world, as stated by Cohen et al. (2007).

Ontologically, research underpinned by an interpretive approach is informed by a constructivist view of the social world (Wahyuni, 2012). Wahyuni summarised the nature of social reality from a constructivist point of view as being socially constructed, subjective, changeable, and multiple. This view of truth echoes the research approach regarding focusing on the assessors as the primary source of understanding and building of their AL. This constructed reality about their AL is not static pre-decided knowledge that is ready for the researcher to discover, judge, or evaluate. This reality is multileveled, complex, and dynamic socially constructed one that is continuously being shaped and is always in an active state because its constituents are changeable and evolving. Different interrelated variables inform it; it is informed by the assessors' technical assessment knowledge, skills, experience, preferences, beliefs, conceptions, socio-contextual domains, educational policies, educational purpose, and students' needs. Therefore, the reality about TAL can only be constructed in its natural settings (Richards, 2003) by its subject or social actors (Howell, 2013). In other words, fulfilling the research purpose and answering the research questions can be derived from the point of view of those who live it, from the emic (insider) point of view through their mind (perception) and their experiences (Howell, 2013). Accordingly, the constructed reality about TAL will be pluralistic, contextualized, and relative because individuals perceive reality differently.

Epistemologically, in research guided by an interpretive research approach, knowledge is generated through a continuous interaction between the

subjects and the social phenomenon to create their meanings (Burns, 2000). The way TAL knowledge was generated, interpreted, and developed by the assessors reflects the same adopted constructivist view. The participants were not passive actors throughout the research study. I involved them in different interactive research activities. They did not only describe their knowledge base of AL; they also shared their views and beliefs about it and their contextual assessment philosophies. The research also encouraged the assessors to challenge the norms, values, and rituals that were practised on them socially and historically by making them reflect on both to free themselves from knowledge imposed on them and make them create a transformative understanding of AL. This critical view gave another lens through which I looked at reality. Participation in the current research study would free the assessors from their taken-for-granted beliefs that were rooted in them through "hegemony", a term that Howell (2013) uses to describe the power exercised on people through consent rather than force. However, it is worth mentioning that this research does not require the assessors to take steps and change this transformative knowledge into transformative practices, test them, and evaluate them like action research. It just aimed to empower and free them by raising their awareness and reflective abilities regarding their AL and practices (Richards, 2003).

### **3.2 Research Methodology**

In the current study, I adopted exploratory practice principles suggested by Allwright (2005), where the emphasis is on understanding and development of the studied phenomenon in its context and by its actors. I explored the assessors' AL with the aim of understanding and development by engaging them in a professional dialogue. This practice assumingly led to integrating newly developed knowledge into existing knowledge of AL and consequently, development in APs (Allwright, 2005). In exploratory methods, the relation between the researcher and the participants is framed by collegiality. Knowledge is generated with the help of the participants (Eisner, 1991). It is a "we research our practice", not "I research others" (Perpignan, 2001, 2003). This relation reflects the relation between the assessors in the current study and me. This relation enhances teachers' image as "knowledge generators" not "knowledge

implementers", as teachers contribution to knowledge base were often undervalued (Locastro, 2000).

I adopted exploratory practices in the current study for different reasons. First, teachers who research their teaching context report significant changes in their understanding (Richards & Farrell, 2015) and experience decrease in the feeling of frustration and isolation (Goswami & Stillman, 1987). Second, exploratory research can also raise the status of the teaching profession in society and most crucially produce knowledge that is useful to teachers, policymakers, academic researchers and teacher educators (Brown & Coombe (2016). Third, it can help teachers find their voice (Rainey, 2000). On the institutions level, an exploratory practice can create a positive, supportive culture that encourages reflection and experimentation (Francis, Hirsch, & Rowland, 1994). To sum up, teachers feel convinced by research that is specific, contextualized, observable, and testable (Borg, 2003); one that is relevant to their needs, problems, challenges. They feel convinced because they are the owners of the research and the research outcomes come from them.

Moreover, I selected the exploratory practice as the research methodology because its characteristics resonate with the research aim and paradigm described in the previous section. First, the exploratory approach is concerned with the subjective opinions and beliefs, which the participants share about the studied phenomena (Punch, 2005). This aspect describes the situation in the current study, as the research aims to understand the assessors' AL from their perspective. The participants are not only required to respond to pre-determined criteria of what constitutes their AL. They are encouraged to bring meaning to AL components by revealing their views about them, their relevance to their assessment contexts, and their interpretation of those components as per their conceptions and their contextual domains.

Second, exploratory practice takes place in the natural setting of the studied phenomena, and it does not attempt to manipulate the situation. It also requires prolonged contact with the environment (Dornyei, 2007). This quality applies to the current study since I conducted this study as a practitioner working cooperatively in the studied context with the assessors. I tried to explore and

seek new uncharted information about their formal/informal assessment preparation, current assessment roles, their confidence with the proposed assessment knowledge base and their beliefs about their context-bound assessment practices. This target goal required me to spend a prolonged time with the participants, which reached up to four months of data collection in the described context. The number of participants in the current study was small (eight), which reflects another aspect of this research samples as being small and purposively selected (Brown & Coombe, 2016).

Finally, in exploratory practice, the researcher tries not to contaminate the emergent nature of the data (Dornyei, 2007) by adopting a minimal biased interference role. This aspect was also implemented in the current study, as I avoided imposing any view or revealing any beliefs during the data collection phase and tried to focus merely on natural emergence of data without directing or influencing the participants. My subjective opinions appeared only during the analysis and interpretation phases.

### **3.3 Data Collection Methods**

I chose self-reporting as a method for exploring the assessors' conceptions, beliefs, views, and reflection. In self-reporting, "respondents are asked to report directly on their behaviours, beliefs, attitudes, or intentions" (Holbrook, 2008:2), unlike other research methods, where the participants are subjected to observation or controlled experiments. In self-reporting, the participants are aware "that their use is based on the assumptions that they can answer the questions posed to them, and they are willing to do so" (Holbrook, 2008:2). The current study used three types of self-reporting methods: direct self-reporting, indirect self-reporting, and open-ended self-description.

First, in direct self-reporting, "the participants are faced with face-valid labels of the constructs and asked to give a summary self-appraisal" (Paulhus & Vazire, 2007: 225). In the current study, the participants directly self-reported two times. The first time they directly self-reported their confidence with and need for AL knowledge base through being exposed to face-valid labels of the AL knowledge constructs in the form of a checklist, to which they respond with

justification during a face to face structured interview. The second time, they directly reported and shared solutions to conflicts that aroused between their context-bound practices and their beliefs about such ways, which I presented to them in the form of a report, to which they responded in another structured interview. Direct self-reporting is a reliable, valid method, especially if the chosen AL constructs and items provide a clear description of the attribute being assessed to ensure face validity. This feature reflects the case in the current research; as I based the checklist, which I constructed, on recommended constructs by the TALiP framework. The items were varied, grouped, and labelled under their respected constructs to help clarify the construct to the participants to increase the participants' confidence and interpretations of the constructs (Goldberg, 1992; Knowles & Condon, 1999). Meanwhile, the themes (constructs), which I used in the report were also based on the explored literature and the major component of assessment suggested by Brookhart (2011).

Second, I also used indirect self-reporting in current research. Indirect-self reporting, unlike direct self-reporting "obscure the constructs being measured" (Paulhus & Vazire, 2007: 225). Indirect self-reporting methods are designed to minimize direct exposure to the research constructs so that the participants will not fake answers due to fear or embarrassment. The researcher's role is to interpret the participants' beliefs, attitudes from the responses of the questions; however, the participants can make inferences about what is being assessed. In the current study, the participants indirectly reported information twice. First, they responded to an open-ended questionnaire, at which they indirectly shared information about their assessment preparation and assessment roles. Second, the participants indirectly reported information about their context-based assessment philosophies and practices, in addition to their beliefs and views about such philosophies and practices in a semi-structured interview.

Third, I used open-ended self-description in the research to derive the participants' free description of their conceptions and beliefs. "Open-ended self-description allows the participants to use any constructs they wish in describing themselves" (Paulhus & Vazire, 2007: 226). However, the researcher can still ask the participants to focus on specific domains during description to enable him/her to code the derived data and quantify it systematically. In the current study, I used

open-ended self-description as a research method in an unstructured interview to derive information about the kind of development that happened to the participants as a result of taking part in the research and going through the different stages of the study.

Self-reporting, with its three types, was chosen as a research method for the current study for different reasons. The self-report method provides rich valid information about the participants' views or beliefs because people possess better quality information about themselves and they are more likely to offer accurate self-relevant information (Lucas & Baird, 2006). Second, it is a motivating method for the participants because no one else will be interested in the result of such exploratory practice more than the participants themselves. Another advantage of the self-report method is that it engages the participants' identity (Hogan & Smither, 2001). This feature resonates with the adopted TALiP framework. TALiP framework focuses on assessors' self-identity construction. This stage is crucial because self-perceptions have a substantial impact on how the participants interact with their contexts, as their behaviour will be affected accordingly as a result of new self-realization (McAdams, 2000; Vazire & Gosling, 2004). Finally, the self-report method is a practical research method because it is "efficient and inexpensive" (Paulhus & Vazire, 2007: 228). It requires coordination and cooperation with the participants on the data collection tools, time, and place. In the current study, the participants and I are practitioners working in the same field and context. They participated voluntarily and willingly self-reported their perceptions and views. It is also an efficient method because many constructs were explored at a time.

On the other hand, the credibility of the self-report method is sometimes questioned based on the assumption that not everything people say about themselves is true. Sometimes the participants are motivated by other motives other than accuracy, such as consistency seeking, self-enhancement, self-presentation, self-consciousness, rapport, transference and modelling (Robins & John, 1997). Some participants may provide socially desirable responses, acquiescent responses (constant agreement/disagreement), or extreme responses either to impress through exaggeration, faking, and lying or for self-



deception in the form of self-favouring bias, self-enhancement, defensiveness, or denial (Paulhus, 2002).

To avoid interference of the pre-mentioned motives in the participants' responses, I adopted the following techniques. First, it was stated clearly that the study was meant mainly for mutual understanding and development, not for evaluation or judgement. The participants were informed that their feedback was useful only if their responses were honest because they would benefit and develop through reflecting on them. Both the participants and I were in a learning position while reporting and discussing their assessment perceptions, practices, and views. Second, I sought autonomous responses even when using tools that required direct answers on specific AL constructs. For example, I avoided using words like "know/do not know"; their confidence level with justification for such choices was sought. Therefore, I made it clear that their views about the constructs were more valuable to the research than the amount of knowledge they had about them. They were informed that just reading through the constructs was a learning experience in itself even if they were not confident with them because this would lead to further discussion in following research stages. Finally, the participants were informed that their identity would be protected and kept anonymous during and after the research; the study would not include any information that exposes their identity or their teaching contexts.

### **3.4 Data Collection Tools**

Qualitative data collection tools were used to collect self-reported data from the eight participants about their AL in practice. I adopted a multi-method approach for data collection; I used a range of data collection techniques to answer the research questions and gain in-depth information in the form of explanation, justifications, views, beliefs, and examples for real-life practices (Wahyuni, 2012). In the current research study, I used five research data collection tools to find answers to the research questions: a questionnaire with open-ended questions, an open-ended checklist in a structured interview, open-ended exploratory questions in a semi-structured interview, an open-ended report in a structured interview, and an open discussion in an unstructured interview.

### **3.4.1 A Questionnaire with Open-Ended Questions**

An open-ended questionnaire with open-ended questions was used to collect qualitative data about the assessors' assessment formal/informal preparation and their assessment roles in their teaching context; and therefore, provide answers to the two components of the first research question. Information about the participants' educational preparation would offer a transparent image of the source of their assessment knowledge. This is a piece of valid information, upon which other components of AL would be built. Meanwhile, information about the participants' assessment roles would also verify my assumption about them as being key informants and accurate, reliable source of their own AL and their context-bound APs. If they were not performing essential assessment tasks and involved in crucial assessment decisions, they would have little to share assessment concepts, design, grading, interpretation, faced challenges, and real assessment experience.

An open-ended questionnaire was selected because it could collect thin information in the form of factual data (level of education, occupation, background, age, gender, race), or behavioural data (what the participants are doing or have done in the past). It is used when the researcher does not know the range of possible answers and therefore, cannot provide a pre-prepared response (Dornyei, 2002). This aspect reflects the case with the two components of the first research question, which is after specific factual behavioural data about the participants' assessment backgrounds and roles. The questionnaire was administrated in a face to face meeting to avoid demotivation, misinterpretation, misreading, detailed irrelevant responses, and leaving out questions (Gillham, 2000). Moreover, the participants preferred to respond orally to the questionnaire in an audio-recorded face to face meeting because it was easier for them to talk about their experiences rather than write about it, and they preferred to do it on campus because they did not have time for extra assignment after work.

The questionnaire was written in a straightforward, simple design to motivate the participants to provide relatively truthful responses; it started with a brief title that summarises the required content. The title was followed by a short description of the questionnaire aim (Dornyei, 2002). It was divided into two main

parts. The first part included open-ended questions that sought responses related to the assessors' formal and informal pre-service and in-service preparation, in addition to their effort for professional assessment development. The second part included questions that sought answers about the participants' assessment role and selection criteria. The wording of the questions was clear, precise, and did not include jargon, ambiguous or complicated terms that the participants might not be familiar with (Dawson, 2007). The questionnaire was short and precise; I had examples of the required data to clarify the range of needed data. The layout of the questionnaire directed the participants on the amount of necessary information. In some questions, they were asked to respond with yes/no; however, they were asked to provide reasons for their answers (Gillham, 2005). For a copy of the questionnaire, see appendix one.

### **3.4.2 A Structured Interview with an Open-Ended Checklist**

An open-ended checklist was used to elicit verbal qualitative data about the participants' confidence level (high, medium, low) with and need for the assessment knowledge base suggested by the TALiP model. The participants were also requested to justify their choices. The collected data from the checklist would give answers to the second research question with its two components.

The checklist was designed in a way that reflects the components of the first level of TALiP framework (Assessment knowledge Base). It is divided into seven main parts. Each part addresses one assessment knowledge component (Disciplinary Knowledge and Pedagogical Content Knowledge, Knowledge of Assessment Purpose, Content, and Methods, Knowledge of Assessment Grading, Knowledge of Feedback, Knowledge of Peer & Self Assessments, Knowledge of Assessment Interpretation & Communication, Knowledge of Assessment Ethics). Each part involved sub-items that represents the sub-knowledge/skills that constitute each knowledge component. The participants were required to choose a confidence level (high, medium, low) in addition to verbally justifying or commenting on their choice. They were also required to comment on the items' relevance to their contexts' practices by choosing (needed or not needed); in addition to justifying their choices. Figure 3.1 summarises main and sub-items of the assessment knowledge base as presented to the participants in the checklist.

## (1) Disciplinary Knowledge & Pedagogical Content Knowledge

- \* Knowledge of Second language Learning Theories
- \* Knowledge of Second Language Teaching Theories
- \* Knowledge of Learning Strategies
- \* Knowledge of Learning Styles
- \* Knowledge of Theories of Language Use (Language Models)

## (2) Knowledge of Assessment Purposes, Content, and Design

### Theoretical

- \* Knowledge of language testing and assessments history
- \* Knowledge of language testing and assessment philosophies
- \* Knowledge of the Impact of adopting testing or Assessment Philosophy
- \* Knowledge of Major Assessment Qualities: Authenticity, Interactiveness, Practicality, Washback, Fairness, Ethicality
- \* Knowledge of Assessment Types (Traditional/Alternative) (Norm- /Criterion-Referenced) (Summative/Formative) (Externally /Internally Mandated) )Standardized/Classroom)
- \* Knowledge of Assessment Methods: (Tests, Portfolios, Performance Assessment, Self-assessment, Peer assessment, Checklists, Interviews, Questionnaires, Role-plays)
- \* Knowledge of Major steps in Language Test Development : (Test Purpose, Construct Definition, Content Specification, test Specification, Item writing)
- \* Knowledge of Cognition Taxonomies Required for Different test items or Tasks (Bloom's / Webb's)

### Practical

- \* Identifying and stating language assessment Purpose
- \* Writing goals and objectives of instructions and consequently their assessments
- \* Aligning curriculum objectives, instruction, and assessment
- \* Defining the language construct(s) an assessment will give information about
- \* Utilizing alternative means of assessments to make decisions based on substantive information
- \* Constructing test specifications/blueprints to design parallel forms of a test
- \* Writing selected-response items such as multiple choice-, true-false, and matching
- \* Designing constructed-response items (for speaking and writing)
- \* Constructing well laid-out and perfectly legible items/tasks
- \* Providing clear explicit unambiguous instructions
- \* Providing examples to make candidates familiar with assessment techniques
- \* Including many items to enhance scoring reliability
- \* Writing test syllabuses to inform test users of test formats where applicable
- \* Designing assessments that are valid not only in terms of course content but also course tasks
- \* Designing assessments that are reliable, authentic, fair, ethical, practical and interactive
- \* Incorporating technologies in assessing students

### (3) Knowledge of Assessment Grading

#### Theoretical

- \* Knowledge of grading rational (norm-, criterion-referenced)
- \* Knowledge of methods that change scores into meaningful composites (points, percent, grades, proficiency levels).
- \* Knowledge of basic linear scaling, weighting scaling, weighting components, and precision of results).
- \* Knowledge of scoring techniques for objectively-marked testing
- \* Knowledge of the nature, purpose, and design of scoring rubrics for subjectively marked tests
- \* Knowledge of consistency and moderation principles for judgment-based assessments.

#### Practical

- \* Constructing scoring schemes that quantify learners' performance into useful information for decisions about learners, classrooms, schools and districts
- \* Designing scoring keys with explicit instructions for closed-ended responses (right/wrong, checklist, multipoint methods including rubrics and rating scales).
- \* Designing scoring rubrics with explicit instructions for open-ended responses (holistic, analytic, primary trait scoring)
- \* Calculating consistency of scores/ standard error of measurement to assess scoring reliability for both closed-ended and open-ended responses
- \* Calculating scorer reliability through scorer reliability co-efficient to quantify the level of agreement among scorers
- \* Designing training workshops for scorers on acceptable approved responses

### (4) Knowledge of Assessment Feedback

#### Theoretical

- \* Knowledge of purposes and principles of feedback
- \* Knowledge of different types of feedback (descriptive, evaluative, supportive) with their respective functions
- \* Knowledge of the strengths and weaknesses of various kinds of feedback (task, process, metacognitive, and self-oriented) in facilitating student learning

#### Practical

- \* Setting educational goals and constructing tasks that will move learners towards these goals
- \* Communicating to learners the criteria for success through different communication modes like telling, showing, and having learners discover
- \* Showing learners the gap between their current stage of development and the next/final stage of development
- \* Intervening in the learning process either by taking an "interventionist" or "interactionist" role
- \* Scaffolding their learning through different mediating exercises "graduated prompt", "testing the limits", and "mediated learning experience".
- \* Using internet resources such as online tutorials and adapt contents to address students' particular needs
- \* Using assessment methods to monitor, learning by collecting formal data (tests) and informal data (observing) of students' language development
- \* Providing encouraging feedback that is meant to modify the learning process not just a score
- \* Monitoring, recording, and reporting student language development
- \* Evaluating how well learners have succeeded in achieving the final learning goals
- \* Improving instruction based on assessment results and feedback

## (5) Knowledge of Peers- and Self-Assessment

### Theoretical

- \* Knowledge of benefits of involving students in assessment
- \* Knowledge of strategies of using self- and peer assessment in different assessment tasks

### Practical

- \* Training learners to effectively participate in assessment
- \* Using self- and peer-assessments to promote learners' ability to notice their learning problems and gaps in their target learning goals
- \* Providing learners with criteria used in judging their work and make them use it to self or peer assess their work and their peers work
- \* Encouraging learners to produce their own rating criteria in groups and use them to peer assess one another.
- \* Training learners to use portfolios, diaries, a continuous cards, digital audios, video diaries, online blogs where learners are expected to collect samples of work and commentary to monitor their own learning process and keep records of how their work improves and develops

## (6) Knowledge of Assessment Interpretation and Communication

### Theoretical

- \* Interpreting data related to test design such as item difficulty and item discrimination
- \* Interpreting concept of reliability in language assessment:  
(Dependability, Item analysis, Reliability Threat)
- \* Interpreting validity in language assessments  
(Construct, Content, Criterion, Consequential Validity, and Validity as argument)
- \* Interpreting data from large-scale tests, namely means, modes, medians, bell curves and has the ability to calculate them
- \* Inferring students' strength and weakness based on collected data to communicate it to different stakeholders

### Practical

- \* Calculating item difficulty and item discrimination for close-ended items to obtain reliable scores
- \* Investigating facility and discrimination indices statistically
- \* Calculating test/item reliability, calculating inter/intra-rater reliability)
- \* Investigating validity using statistical procedures
- \* Articulating interpretation of norm- and criterion- referenced assessment results to a variety of audience: student, school, directors
- \* Communicating interpretations using different reporting tools: Report cards, documents, criteria, guidance counselling.
- \* Using software such as Statistical Package to communicate results
- \* Participating in committee or school-wide discussions about reforms in assessment related issues, curriculum, materials, grading policies, accountability policies, school evaluation and school evaluation.
- \* Having the skills to coach students to analyze their own assessment results, track their own learning, communicate about their own learning, and plan next steps.

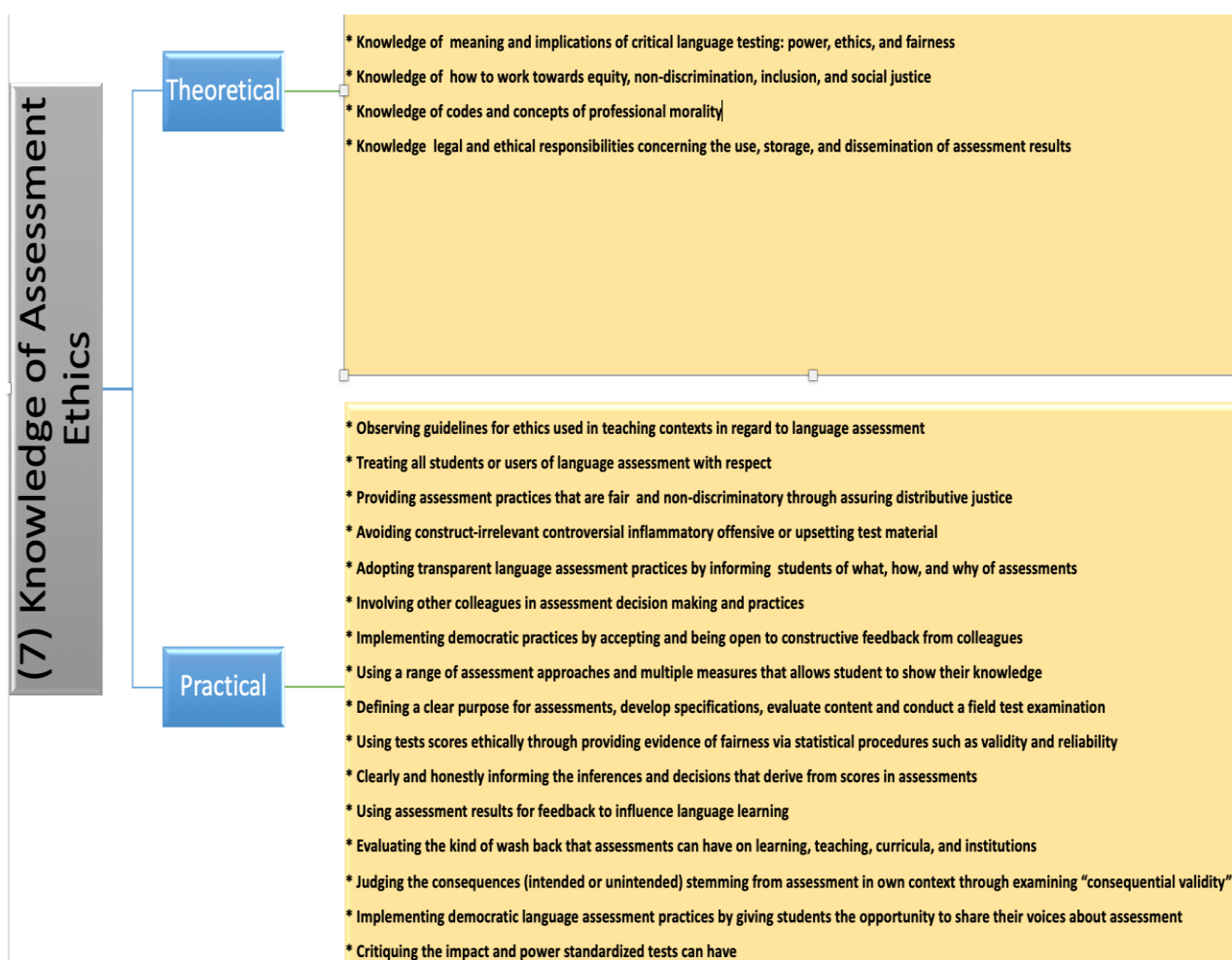


Figure 3.1. Assessment Knowledge Base as Presented in the Checklist.

The sub-items of each component were derived from the explored literature and adopted from established proper tools. Many of the items/sub-constructs describing each knowledge component were adopted from Giraldo's (2018) core list of assessment knowledge and explored literature in chapter two (Brookhart, 2011; Inbar-Lourie, 2008, 2013; Fulcher, 2010, 2012; Fulcher & Davidson, 2007; Popham, 2009, 2011; Taylor, 2013). Different aspects were taken into consideration concerning items wording and phrasing. Words that are complicated, technical or jargonized were avoided. The items were clarified with examples and phrases. Items that were negative, ambiguous, personal, double-barrelled were avoided (Dornyei, 2002). For a copy of the checklist, see appendix two.

The checklist was administrated in a face to face audio-recorded structured interview with the same eight participants. Structured interviews represent data collection in its controlled form (Heigham & Crocker, 2009). It is

used when the researcher is after specific information. The collected data can be easily compared across the participants. It is a kind of spoken questionnaire; however, it overcomes most of the pitfalls of questionnaires. The researcher can guarantee that the participants do not miss any of the items (Gillham, 2005). Any misunderstanding or misinterpretation of items can be addressed by providing clarification, examples, or paraphrasing of items (Dornyei, 2002). Simplicity and superficiality of answers can be overcome by requiring justification for choices or using open-ended questions (Askey & Knight, 1999). In structured interviews, the researcher uses an interview schedule or a scaling instrument to guide the interview. In the current study, the checklist was used as the interview schedule. One advantage of using the checklist is that I incorporated pre-coded answer blocks, and was able to transfer responses into their coded form (Askey & Knight, 1999: 90). Using Coded answers blocks saved time during the later stages of the research because it was easier to group the responses or codes under their respective themes.

A structured interview with an open-ended checklist was chosen to collect data related to the second research questions with its two components for different reasons. First, the second research question is after data that describe the participants' confidence with the AL knowledge base suggested by the TALiP framework and their beliefs regarding their need for such knowledge. Therefore, they need to be faced with the AL theoretical constructs and its sub-skills to interpret them, provide answers about their confidence level with such theoretical knowledge, provide their views about such knowledge, and finally judge their relevance to their assessment context. Second, exposing the participants to the primary theoretical constructs and sub-skills of AL is a learning experience in itself. It is also an opportunity for development for those who have not been exposed to assessment academic assessment preparation, as the current exploratory study does not only aim to understand the participants' AL, but it tries to contribute in conceptualizing their AL, and consequently reconstruct their assessment identities as assessors. Finally, such negotiation about the proposed theoretical base knowledge acted as a solid base for more advanced levels of AL, where they were required to reflect on their practices, context-bound philosophies, and compare them to theories to come up with solutions to uncertainties that appeared during the research phases.



### **3.4.3 A Semi-Structured Interview with Open-Ended Questions**

An audio-recorded semi-structured interview with open-ended questions was conducted with the eight participants to elicit information about their context-adopted APs and their beliefs about these practices. Data collected from the interview would provide answers to the third research question with its two components.

Before constructing the interview, I created an interview guide (interview questions) (Arksey & Knight, 1999). The guide was derived from the explored literature and AL theoretical knowledge base suggested by the TALiP model. The questions tried to elicit information about the context-adopted assessment purpose, source, methods, decision-makers, design process, administration, grading, feedback, the inclusion of peer & self-assessment, communication of results, and ethical procedures, in addition to the participants' views about these discussed practices. For a copy of the interview questions, see appendix three. The questions were open-ended, contextual, exploratory, and generative to collect thick descriptive qualitative in-depth data, which is another crucial feature of exploratory studies (Holliday, 2001). Dornyei (2002) mentioned that open-ended questions could be used to collect qualitative data if they include attitudinal questions that were concerned with beliefs and values. The selected questions were unambiguous and straightforward, and I avoided using double-barrelled questions. All questions aimed to elicit data with very little interference, except for more clarification or elaboration.

A semi-structured interview was chosen because it is a suitable technique for collecting detailed data related to attitude and beliefs, which are the main focus of the two components of the third research question. It can also allow flexibility to let the participants lead (Silverman, 2001, p. 51), and therefore provide the chance for personal interaction and eliciting more in-depth knowledge (Arksey & Knight, 1999; Heigham & Croker, 2009). Semi-structured interviews are designed to give respondents the chance to express ideas and are thus "appropriate for eliciting teachers' perspectives and views" (Arksey & knight, 1999:96). "They hold the possibility of understanding the lived world from the perspective of the participants involved" (Heigham & Croker, 2009:187). Moreover, interviews provide the option of asking for clarification, which can

make the relationship between the interviewer and the interviewee collaborative and exploratory (Richards 2003). They can also provide other nonverbal data elicited from the participants' tone, facial expressions, and body language, as mentioned by Richards (2003).

#### **3.4.4 A Structured Interview with an Open-Ended Report**

An open-ended report was used to collect data on how the participants negotiate between their views about assessments and their context-adopted APs. One report template was designed for each participant. I divided the report template into themes that reflected the assessment components explored in the previous semi-structured interview (purpose, source, methods, decision-makers, design process, administration, grading, feedback, the inclusion of peer & self-assessment, communication of results, and ethical procedures). The participants' responses to the previous interview were transcribed, organised, and copied in each report. Three columns were added in front of each theme. In the first column, I added the participant's description of the context-adopted practices for each assessment component. In the second column, I added the participant's views about the adopted assessment practices. The third column was left empty for each participant to add solutions to detected discrepancies during the structured interview. The report aimed to collect compromises and solutions to the detected problems and provide answers to the fourth research question. For a copy of the report template, see appendix four.

Each participant responded to the open-ended report in an audio-recorded structured interview. I sent each participant his/her report template two days before the interview to have the chance to read through it, notice the discrepancies, critically reflect on them, and think of appropriate solutions and compromises. During the interview, I gave the participant the chance to share any dissatisfaction or disagreement with the report content. Each participant provided his/her critical views and suggested solutions to noticed discrepancies. After the structured interview, the participant's responses were transcribed and added to the third column to finalise a whole complete report for each participant. For a sample of a completed report template, see appendix 10.5.

The open-ended report was used as a tool for collecting data for two reasons. First, it was used as “member check” technique for the collected data

from the semi-structured interview by providing the participants with reports based on transcription and analysis of their interviews to indicate their agreement or disagreement with the collected data. So, it is used as a kind of validation for the content of the analysis (Wahyuni, 2012), which could contribute to enhancing the research credibility. Second, the report was used to pave the way for the last level in exploring, understanding and developing the participants' AL, which is the compromises made between their beliefs and real contextual practices. The participants were put in a problem-solving situation, where they were required to reflect on assessment problems they face in their contexts critically; the participants were expected to come up with solutions to conflicts they meet in real life. This critical reflection was an essential part in the study because it involved understanding, enlightenment, and development that could have happened to both the interviewer and the interviewee, which is one main aim of the research study.

#### **3.4.5 An Unstructured Interview**

An audio-recorded unstructured interview was used to elicit information on how the participants' AL was developed as a result of going through the TALiP levels and taking part in the current exploratory study, and therefore provide answers to the fifth research question. The participants were asked to talk freely about their experience with the current research and its impact on them. They were also encouraged to talk about their future development and enhancement as a result of taking part in the full research study. The nature of the used questions invited description, explanation, reflection, and interpretation of the research experience (Patton, 2002: 348-351); I tried indirectly to elicit the participants' knowledge, opinions, and feelings about the current research practices to reach a noticing phase, at which the participants can recognize and realise the kind of development that happened to them. For a copy of the brief interview guide, see appendix five.

The unstructured interview was chosen to elicit answers to the fifth research question because it is used to achieve a holistic understanding of the participants' point of view or situation (Dawson, 2007). In such interviews, the researcher is not aware of the details of the data because they are based on the participants' experience with a particular situation. The participants are asked to

talk freely, and the researcher asks a few questions with little directional influence from his/her side. This feature reflects the situation in the last stage of the research because I was after creating a relaxed atmosphere, in which the participants could reveal their own experience and possible development that might have happened to them. I was aware that some unexpected themes would emerge, as each participant would be sharing a different experience (Dornyei, 2007). It was also found suitable to the last stage of the study because after spending four-month period with the participants on the current project, the participants and I had developed a sense of mutual trust and shared interest, which enabled us to talk freely about personal experience (Brown & Coombe, 2016).

The problem with the unstructured interview was that I had to remain focused to probe for more details with little interference. Moreover, the interview resulted in a great deal of data, which were difficult to analyse (Dawson, 2007). However, the richness and importance of the data elicited from the interview were worth spending time focusing on their responses and trying slightly to direct them if they deviated away from the original topic.

### **3.5 Research Procedures**

#### **3.5.1 Sampling Procedures**

The main aim of sampling in the current research was to find participants who can provide "rich and varied insights into the phenomenon under investigation", to maximize what can be learnt; therefore, I believe that this goal is best achieved by "purposeful" sampling (Dornyei, 2007:126). Purposeful sampling is a crucial feature of exploratory interpretive research, in which participants are selected through non-random methods based on their knowledge of essential information that is vital to the research questions (Lodico, Spaulding, & Voegtle, 2006).

In this study, eight experienced assessors, who were involved in assessment decision-making in an EPP in one university in Kuwait were chosen to guarantee that their knowledge is relevant to the topic under investigation, which reflects Patton's (1990) description of the sampled participants as being

"key informants". The participants were considered vital informants because they were responsible for assessment design, method selection, content selection, rubric design, distribution of grading tasks to other teachers, administering assessments, and communication and interpretation of results. Their assessments' decisions had a crucial impact on students' future and continuation in the described context. They were a typical homogenous sample (Dornyei, 2007) that shared a common core experience, which was relevant to the current study. Therefore, I believe that the shared experience of the participants explained the observed problem of assessment in the studied context. It allowed the participants and myself to negotiate meaning concerning assessment philosophies and practices and engage in a professional dialogue that is meant for mutual conceptual development of assessment and assessors' identities.

There were two challenges that I met with the selected sample. One was to persuade them to take part in the study without feeling a threat regarding their shared experience and their personal views that might contradict with their context-bound philosophies and practices. The second was to persuade them to take part in data collection procedures, which required spending quality time with me and being involved in different research tasks in addition to their assessment tasks and daily teaching responsibilities.

To handle the first challenge, once approval was obtained from the university ethics committee for data collection, each participant was contacted verbally. For copies of the ethics application form and the ethics approval certificate, see appendix six and seven. A brief verbal description of the study was shared with each participant regarding the objective, duration, and data collection tools. They were informed that their identity and the programme details would be kept anonymous throughout the research and after finalizing the research project. I confirmed that I would not include in the study information that might contribute in revealing their identities (name, age, nationality); only data related to education level and years of experience would be used to support their status in the research as a typical sample for the current study.

They were informed that they would be given a signed information sheet, which outlines the project's aim, their roles, and their rights, in addition to adopted

protection steps to protect their identity, shared data, and safety. They were also informed that they had the right to withdraw from the study at any time, and all shared data would be destroyed. The sheet also would include my contact details, research supervisor, and ethics chairman to contact them if they feel unhappy about any research procedures. I explained that the data would be collected specifically for the research, and the raw data would not be available to anyone else other than the research committee and me if needed. They were assured that none of the output (the dissertation, article, report, conference, seminar, and presentation) would provide information that would identify their names or their teaching contexts.

To handle the second challenge, I explained the goal and the benefit behind taking part in the research project, which was meant for mutual conceptual development as a team working in the same context. I explained briefly the tasks in which they would be involved. The expected participants were briefed that the data would be collected over four months. I acknowledged that I would not rush them to complete takes within a short period. Critical working time would be avoided, especially time; at which they would be performing assessment or grading tasks. It is worth mentioning that the initial number of participants that was targeted was twelve; however, only eight agreed to take part in the study after sharing the pre-mentioned information. A brief description of the participants' background is provided in table 3.1.

<b>Pseudonyms</b>	<b>Gender</b>	<b>Qualification</b>	<b>Experience with Preparatory Programs</b>
Sherry	Female	PhD in Linguistics	More than ten years
Janset	Female	Master's in Literature	More than 20 years
Tok	Male	Master's in Education	More than ten years
Rawan	Female	Master's in Education	More than 20 years
Mirjana	Female	Master's in Linguistics	More than ten years
Jasmine	Female	Master in Linguistics	Less than ten years
Rose	Female	Masters' in TEFL	More than ten years
Talen	Female	Bachelors' in English	Less than ten years

*Table 3.1 The Participants' Pseudonyms, Gender, Qualification and Experience*

### **3.5.2 Data Collection Procedures**

Upon receiving verbal approvals from the participants, an information sheet and a consent form were prepared for each participant (see appendix eight and nine). Before conducting any data collecting procedure, a brief meeting was conducted with each participant to explain the five steps of data collection procedures and the purpose behind each step. They were informed that data collection would require five face to face meetings, one every month, each would last for a maximum of an hour. Time and location of the meetings would be chosen as per the participant's preference, availability, and convenient time. Each participant was asked to decide on a convenient time and place for conducting the interviews (Arksey & Knight, 1999). Accordingly, an interview appointment schedule was created and shared with the participants; they were informed that a reminder of each appointment would be sent two days ahead of each meeting. For any inconvenience, time would be rescheduled as per their availability.

During the meeting, the participants and I agreed on audio-recording the interviews. Audio recording helps the researcher to focus on what is said. It also records tone, voice, emphasis, and pauses, which add to the researcher's understanding and interpretation during the transcription and analysis phase (Arksey & Knight, 1999). It makes the participants feel that their shared information is taken seriously and treated professionally. I used English as the language for the interview because this was the common language between the participants and me. At the end of each meeting, each participant was given the consent form to sign and the information sheet, which outlines the research title, purpose, steps, and conditions.

Before conducting the interviews, they were piloted with an experienced assessor working in the same context. A meeting was conducted with the pilot interviewee at the end of each phase to record feedback on the interview's length, tool, questions, clarity, attitude, interaction, disruptive behaviour, and physical conditions of the interview. After the first meeting, the pilot commented on the open-ended questionnaire; she mentioned that questions were requiring the same responses (assessment role, and assessment task); she recommended keeping only one of them. The pilot recommendation was addressed, and this

section was adjusted. She also mentioned that the rest of the questions were clear and straight forward.

After the second meeting, where she was exposed to the detailed checklist, she had additional comments. She found many of the items difficult to understand because they included terminology, with which she was not familiar. She recommended providing a short brief description of each item. She also mentioned that there were items, which were redundant and repeated in more than one section. She was given the checklist, and she highlighted them. She found it challenging to finish the checklist in one hour within one interview because it was lengthy and very detailed. Accordingly, the checklist was revised against redundancy, and a brief explanation of every item was provided under it. She did not face any issues with the other three tools. Only in the unstructured interview, she did not know what to say at the very beginning when she was asked to comment on her participation. She was directed with few clues, such as reminding her of the different research stages. The feedback collected from the pilot interviewee developed and improved the data collection tools before being used. (Richard, 2003).

On every scheduled meeting with the main participants, I arrived on time with notes and prints of the data collection tools. Recording tool was checked in advance. I started by giving a brief description of regular interview practices, in addition to a word of thanks. I also asked for permission to audiotape the interview and provide space for any questions from the interviewee' side before starting the interview (Brown & Coombe, 2016; Dornyei, 2000). During the interviews, I tried to keep control of the interview and at the same time, be very flexible with the order of the answers, with minimal interruptions to maintain the natural flow of conversation. I avoided controlling questions that require brief responses. To assure mutual understanding throughout the interviews, I offered clarification, welcomed questions, paraphrased queries, and summarised what was understood after each lengthy conversation.

Throughout the data collection phase, I was a careful critical active listener through watching emotional tone, body language, facial expressions and looking for keywords that give perspectives, explanation, and experience. This technique



helped in introducing new topics and questions that were opened up by the participant, which were not on the interview guide. I gave encouraging feedback and comforting gestures that showed interest in the participants' ideas; this inspired the participants to reveal more ideas (Arksey & Knight, 1999). All interviews ended by expressing gratitude to the participants' contribution and valuable shared data.

### **3.6 Data Analysis**

Data analysis involves drawing inferences from raw data (Wahyuni, 2012:75). Raw data were collected in the current research from multi-methods (Patton, 2002). The analysis procedures in the current study went through two main steps: data preparation and qualitative data analysis (Boeije, 2010). Data preparation went through three stages. The first step was a neat archive to store the data to be able to retrieve them easily. The five recorded audio files for each participant were stored electronically in a separate folder under his/her pseudonym on my password-protected computer. Second, the audio recorded interviews for each participant were transcribed, and a soft folder was created for each participant, in which the related transcripts were stored (Richards, 2003). The transcriptions were revised against the audio recording to check the transcription accuracy. Linguistic details such as laughter and irrelevant phrases were deleted (Oliver, Serovich, & Mason, 2005) from each transcription. Third, data were cleaned for ethical consideration in term of anonymity and confidentiality; all information that could have identified the participants or their teaching institution was omitted. This step was considered in case the documents were requested by the research supervisor or the university ethic committee. For examples of all transcriptions and completed research tools of one participant, see appendix ten.

Second, the prepared cleaned data were subjected to analysis. The analysis process involved transforming the collected raw data into findings that enabled interpretation (Wahyuni, 2012). Content analysis was adopted in analysing the five collected texts for each participant through identifying patterns and themes within data, which is referred to as thematic analysis (Given, 2008). For each research tool, the eight generated texts were read, compared, and

common themes were coded. The coded themes were used as the core units for comparing data across the eight texts generated from each tool (Thomas, 2003). The research tools were already divided into themes and sub-themes, which facilitated comparing the collected data across the eight texts generated from each research tool. Accordingly, the data derived from the five research tools were already grouped under specific themes guided by the components and subcomponents set in each research tool.

### **3.6.1 Analysis Method Conducted on Data Collected from the Open-Ended Questionnaire**

Data collected from each questionnaire were transcribed in a word document, read, and analysed after being thematically categorized and coded under two main themes: Theme A (educational & practical preparation) and Theme B (involvement in assessment tasks). These two themes correspond with the two main components of the first research questions and the two main parts of the open-ended questionnaire.

Data related to Theme A were sub-categorized and coded under two main sub-themes: Theme A.1 (pre-service preparation) and Theme A.2 (in-service preparation). These two sub-categories correspond with section one in the questionnaire, which focuses on teachers' assessment preparation. Theme A.1 was further subcategorized into theme A.1.1 (theoretical preparation) and Theme A.1.2 (practical preparation). Theme A.2 was subcategorized into theme A.2.1 (theoretical preparation), Theme A.2.2 (practical preparation), and Theme A.2.3 (teachers' effort).

Data related to Theme B were sub-categorized and coded under two main sub-themes: Theme B.1 (precise assessment roles) and Theme B.2 (selection criteria). Data categorized under these two sub-themes corresponds with the answers of the two questions listed under section two in the open-ended questionnaire, which focuses on teachers' assessment experience. For an illustration of thematic categorization and coding of data, see Figure 3.2

### Thematic Coding of Data Collected from the Open-Ended Questionnaire

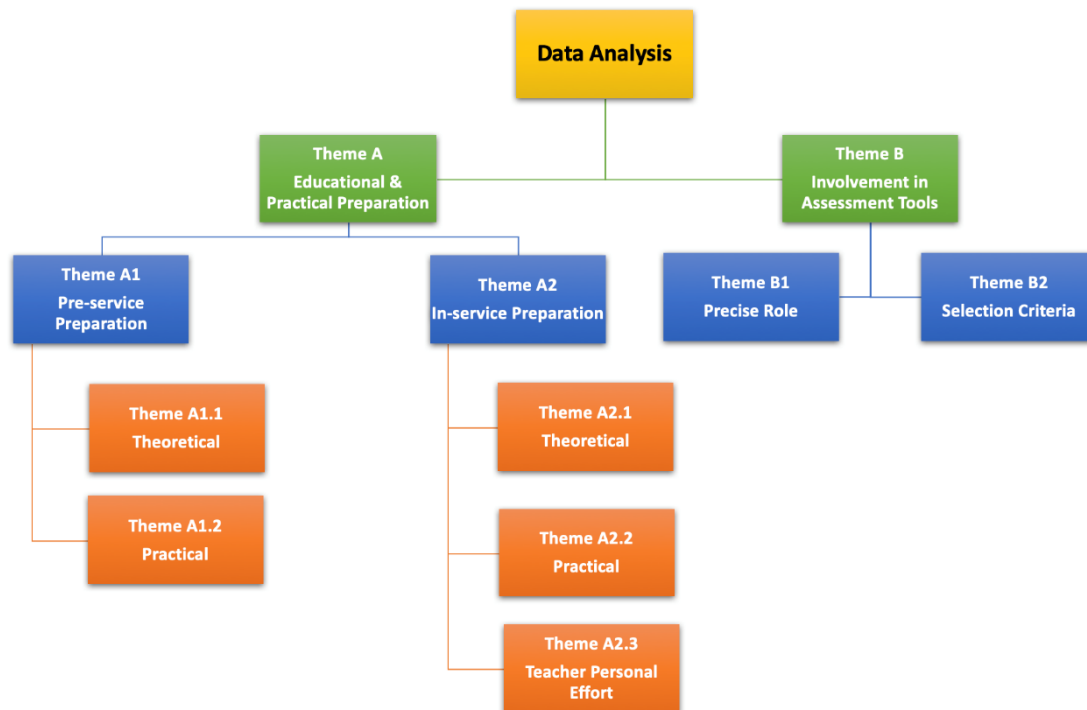


Figure 3.2. Thematic Categorization and Coding of Data Collected from the Questionnaire.

After coding data in each transcribed document, a comparison table was designed; the table consisted of eight columns each carrying the pseudonym of each participant and seven rows, each bearing the title of a sub-themes (A.1.1; A.1.2, A.2.1, A.2.2, A.2.3, B.1, B.2). The coded data in each transcribed were listed and grouped in the table in front of its concerned sub-theme and under the name of the concerned participant as specified in table 3.1, appendix eleven.

### 3.6.2 Analysis Method Conducted on Data Collected from the Open-Ended Checklist

The checklist (used in collecting data for questions 2.A and 2.B) was already thematically categorized and coded into seven themes addressing the seven components of assessment knowledge base suggested by the TALiP model. Each theme was subcategorized into theoretical knowledge and practical knowledge, under which sub-skills were listed (except component one, which had only theoretical knowledge).

A soft copy of the checklist was created for each participant. The written ticks they added on the hard copy were copied into the soft document kept for

each participant. The audio-recorded reasons each participant gave for his/her choices were transcribed and typed into the reason slot of the soft copy. At the end of the transcription and typing process, the researcher had one complete soft copy of the checklist for each participant, which included their ticks (choices) and the reason they gave for such selection.



*Figure 3.3. Thematic Categorization and Coding of Data Collected from the Checklist.*

To compare the participants' responses about their confidence with and need for assessment knowledge components and sub-components, I compiled the choices of the eight participants and the reasons they gave for their choices into two tables, one for their confidence levels and the other one for their needs for such knowledge. Table one (confidence level and reasons) was divided into seven sections; each section was about one component of the seven competencies (seven themes). Under each component, rows were created, each carrying the title of the sub-component of the concerted assessment component. In front of each component, there were eight columns, each bearing the pseudonym of each participant. The participants' responses were typed under their pseudonyms in front of the sub-skill, about which they gave their responses.

At the end of the coping process, one complete table was created that included the answers the eight participants provided for each component and sub-component of the assessment knowledge next to one another ready for comparison and dissemination of the finding related to each component. The same process was followed in table two (the need for each competence). For copies of table 3.2.and table 3.3, see appendix twelve and thirteen.

### **3.6.3 Analysis Method Conducted on Data Collected from the Semi-Structured Interview**

Data collected from the interviews were transcribed directly from the audios in word documents, one for each participant. Data were thematically categorized into thirteen themes, which corresponds assessment features discussed in the semi-structured interview. Themes were coded from 1 to 13. Each code corresponds to one component of assessment as specified in the below diagram. Each code provided for each theme was sub-coded into two codes, for example, theme 1 was sub coded into 1.1 & 1.2, where 1.1 represents the participant's response regarding how this theme (assessment feature) is practised in his/her context, and 1.2 presents the participant's view about this theme (assessment feature).

Each word document was coded accordingly with 26 codes (13 for the actual practices and 13 for the believed about those practices). After each document was coded, two tables were created, one for the context-adopted assessment practices and one for assessors' beliefs about those practices. Each table consisted of thirteen rows carrying the codes of the themes and eight columns, each bearing the pseudonym of the participants. Answers related to each theme were grouped in front of it and under its concerned pseudonym. Categorizing and grouping of data in tables enabled me to allocate the participants responses related to each theme beside one another ready for comparison and dissemination of the findings. For a copies of the two analyses tables, see tables 3.4 & 3.5, appendix fourteen and fifteen.

## Thematic Categorizing and Coding of Assessment Features



*Figure 3.4. Thematic Categorization and Coding of Data Collected from the Semi-Structured Interview*

### 3.6.4 Analysis Method Conducted on Data Collected from the Report

The audio recorded data for each participant were transcribed in a separate word document. The data in the word document were categorized into thirteen themes equivalent to the ones in the original report and specified in figure 3.3. Each theme was coded with a number; there were 13 codes in each document numbered from 1-13. The thirteen categorized, coded data were transferred from each participant's word document to a soft copy of his/her report, where it was added to the third empty column of each report; each coded data was added in front of its equivalent theme number. By the end, each participant

had a complete report with three columns, one for what he/she reported about his/her context-adopted assessment component (theme), one for his/her belief about each component, and a third for a solution he/she reported for each assessment component.

To group responses of all participants to each theme for comparison, a table was created, which consisted of thirteen rows carrying the codes of the thirteen pre-categorized themes and eight columns, each bearing the pseudonym of the participants. Data related to each theme were grouped in front of it and under its concerned pseudonym. This set up enabled me to allocate the participants responses related to each theme beside one another, ready for comparison and dissemination of the findings. For a copy of the analyses table, see table 3.6, appendix sixteen.

### **3.6.5 Analysis Method Conducted on Data Collected from the Un-Structured Interview**

The audio recorded data collected during the unstructured interviews were transcribed directly in word documents, one for each participant. All word documents were read thoroughly to find common themes for comparison among the eight documents. Six themes were detected in the eight transcriptions: (A) Impact of taking part in the open-ended questionnaire, (B) Impact of going through the detailed checklist, (C) Impact of taking part in the semi-structured interview with the open-ended questions (D) Impact of taking part in the structured interview, which required report filling, (E) Impact of taking part in the research, and (F) Impact of participation in the study on future decisions related to assessments. Data were thematically categorized and coded in each transcription. After each document was coded, a table was created, which consisted of six rows carrying the codes of the six themes (A-E) and eight columns, each bearing the pseudonym of the participants. Data related to each theme were grouped in front of it and under its concerned pseudonym. Categorizing and grouping of data in the table enabled me to allocate the participants responses related to each theme beside one another ready for comparison and dissemination of the findings. For a copy of the analyses table, see table 3.7, appendix seventeen.

### Thematic Categorizing and Coding of Data Collected from Unstructured Interview

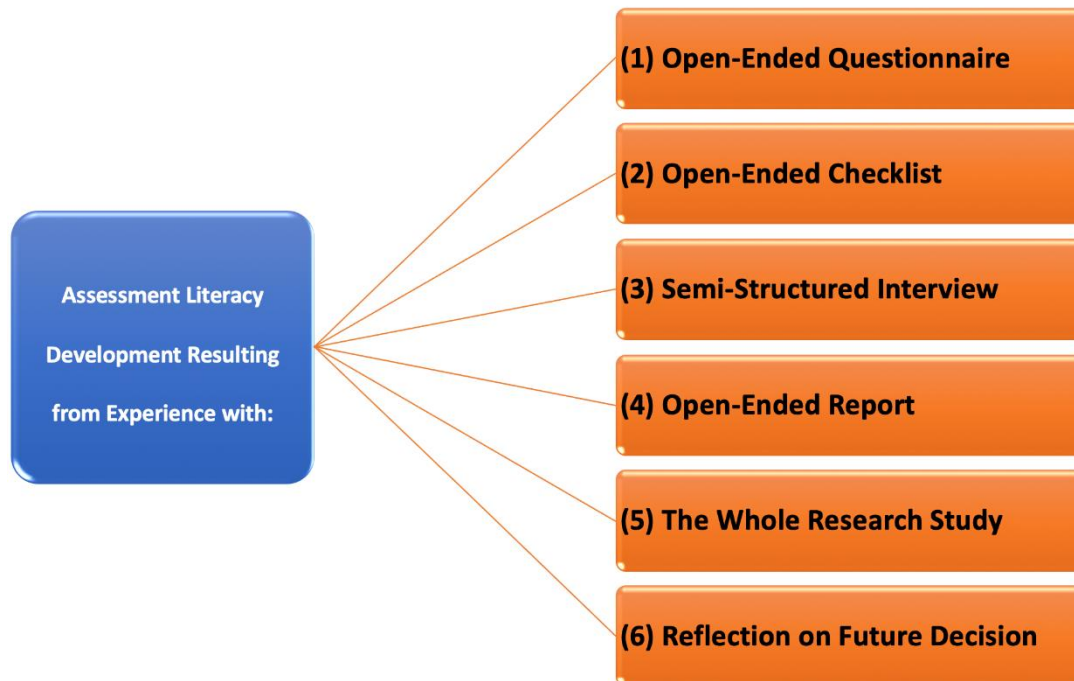


Figure 3.5. Thematic Categorization and Coding of Data Collected from the Unstructured Interviews

## 3.7 Research Trustworthiness

Establishing trustworthiness is central to any conception of quality in qualitative research (Guba & Lincoln, 1989, 1994). Guba and Lincoln developed five criteria to establish the trustworthiness of qualitative research; credibility (presenting convincing results), transferability (possibility of applying findings in other contexts), dependability (findings can endure over time), conformability (minimizing personal bias), and authenticity (creating a sophisticated but temporary consensus of views about what is considered true).

I adopted a set of standards during the study to assure trustworthiness. First, I achieved credibility through thorough engagement in the field, as I collected the data over four months, in which I engaged the participants in different research activities. I approached each participant at least five times in addition to the first introductory meeting. I managed to have access to the data source and participants as a practitioner working in the described contexts. Therefore, I spent considerable time with the participants who self-reported their



conceptions, beliefs, views, and critical reflective ideas about their own AL in practice using the five research tools. The multiple perspectives that the research presented added richness to the findings, especially that I accompanied the proposed findings by a rich, detailed description of the context, participants, and research steps.

Moreover, I used a "member check" technique by providing the participants with an open-ended report based on the transcription of the semi-structured interview to indicate their agreement with the collected data and to reflect on their responses critically. I designed the study in a way that made the participants co-researchers; they knew that the research was after mutual understanding and development. At the end of the study, I shared the final report with "a critical friend/the pilot" working in the same field to respond and provide feedback on the adopted research design, collected data, analysis and reporting findings. I used feedback collected from both the "member checking" and "critical friend" (Rossman & Rallis, 2003) in enhancing data quality, finding dissimulation, and chapter organization.

Second, I tried to adhere to rigorous steps to maintain a high level of transparency. I adopted a conceptual framework to guide the research (TALiP); I explicitly explained the framework and backed it up by theoretical perspectives presented in chapter two. I presented the research approach, adopted paradigm, methodology, and data collection tools along with justification for their choices. I documented all research procedures and steps; this includes data collection phase, analysis phases and dissimulation of finding. I recorded and stored safely all collected data, whether in the form of hard copies or audio-recorded files. The emerging data were not contaminated because I tried throughout the research not to manipulate them by adding /deleting especially during the data collection phase. I tried to report the exact meaning the participants wanted to convey without any bias. Moreover, I provided a detailed description of the context, data collection method, procedures, analysis tools, and steps followed to reach the findings to give the readers sufficient information to be able to judge the applicability and transferability of results to other contexts. I added samples of transcriptions, analysis steps, analysis tables, and quotations in the appendix and findings and discussion chapter.

Third, I tried to achieve dependability through adopting an auditing practice through documenting the data, method, decisions, in addition to adopting a reflexive behaviour, which includes a self-critical account of how the research is done (Atkinson & Delamont, 2010). Auditing also helped in achieving conformability, as it minimized personal bias. Finally, I demonstrated authenticity by achieving fairness by showing a range of different realities about the observed problem. I provided a more sophisticated understanding of the problem in the form of the research finding, which I reached through consensus (ontological authenticity), appreciating the participants' views (educative authenticity), and simulating some form of action through providing recommendations based on the finding of the research (catalytic authenticity) (Guba & Lincoln, 1994).

### **3.8 Ethical Considerations**

Throughout the current study, I followed the recommended ethical research procedures for qualitative research (Alderson and Morrow, 2004). First, my integrity as a researcher was maintained by sharing verbal and official signed information letters and consent forms with the participants assuring that no fabrication, falsification, or misrepresenting of data would take place. I informed the participants that no harm would happen to them as a result of taking part in the study or publishing it (Shohamy, 2004). I did not give a detailed description of the participants or the research context to protect their identity and not to cause them any harm as a result of taking part in the current study. I assured them that they would benefit from taking part in the study by providing a brief explanation of the purpose and procedures of the research. I made it clear from the beginning that I would not use the tools/techniques/questions in the study to evaluate or judge their AL; I would use them to support their knowledge and professional development.

I informed all participants about the means of protecting their privacy, anonymity and security procedures of data storage. The participants' identity was kept anonymous. I used pseudonyms to refer to the participants during the interaction, data recording, transcription, analysis, and writing up. I informed them that the data would be analysed specifically for the research and none of the output (the dissertation, article, report, conference, seminar, presentation) would

provide information that would identify their names or name of the programme, at which they work. I explained that they had the right to refuse to answer any question or withdraw from the study without offering any explanation. For safety and security, I stored all audio recorded data and hard-copy documents that included data provided by the participants in a locked filing cabinet, and I deleted the audio data immediately from the digital recorder after being stored on a password-protected computer. I intend to destroy all data after the end of the study.

I documented all the previous precautions to preserve the right of both the participants and myself. The participants read the terms included in the information letter and consent form before starting the research procedures. The information letter included information about the research purpose, practices, and the reasons for being selected. It also had a description of the participants' tasks in the research, gained benefits, possible risks, and the right to withdraw or ask for clarification. I added the means of protecting the participants' anonymity and storing the provided data. I also included my contact details, in addition to the supervisor's contacts and the research ethics and governance manager contact mail. The consent form included terms regarding the voluntary nature of their participation, access to their provided data, research tools, duration of the research, the anonymity of their identity and their programme, detailed tasks and required participation time in each, and data storage details. The consent form required both the participants' signature and mine as well. Each participant received a copy of the information letter and the consent form; I kept the other documents in the project file.

## **Chapter Four: Findings and Discussion**

### **Introduction**

The current chapter presents the findings and discussion of the eight research questions. It is divided into two main sections. Section one presents the findings of the research questions, and section two presents the discussion of the findings. Section one is divided into eight parts; each part focuses on one research question. Each of the eight parts is sub-divided into sub-parts titled with the themes that resulted from the data analysis related to the concerned question. The findings related to each theme is placed under it. Section two is divided into eight parts. Each part focuses on the discussion of one research question findings. The discussion shows how the answer to each question relates to the research body conducted on AL to support, contradict, or add to it. It also shows how the findings of each question relate to the research purpose and how far the findings contributed to achieving its objectives.

### **4.1 Findings Related to the Research Questions**

#### **4.1.1 Findings Related to Question 1.A**

*Question 1.A How are the assessors prepared for their assessment tasks?*

Question 1.A focuses on eliciting data related to the participants' formal and informal assessment preparation during their pre-service and in-service phases. Answers to this question would provide a clear image of the participants' theoretical and practical assessment preparation, which is a significant primary component of their AL, upon which other components of their AL will be built. Answer to question 1. A was reached after comparing responses provided by the eight participants on three themes related to their educational preparation for their assessment tasks: teachers' pre-service preparation, in-service preparation, and self-development efforts.

##### **4.1.1.1 Pre-Service Preparation**

Six participants emphasised that they have not received any theoretical assessment preparation. The only theoretical preparation they received before starting their teaching career was in the form of teaching courses, which focused

mainly on teaching methodology. Rose, who obtained two degrees in education before starting her teaching career, reported that she had not taken any course during her academic preparation related to assessments. Jasmine added: "Most of the textbooks I was exposed to were related to applied linguistics in general, nothing was about how to test". The only two participants who mentioned receiving theoretical preparation on assessments before starting their teaching career admitted that these courses were concise and not in-depth. Tok explained: "I did not have one course, which focused on testing; it was a component of a methodology course". Seven participants denied taking any practical preparation or training on assessments before starting teaching. Jasmine commented: "I do not remember having any training on how to create an exam before starting my assessment task". Only Janset mentioned taking a small-scale training session during her teaching preparation phase on how to create small exams tasks.

#### **4.1.1.2 In-Service Preparation**

The eight participants denied receiving any theoretical guidance on assessments in the form of lectures, seminars, or tutorials in the described context. They also acknowledged that they had not received any practical support in the form of workshops or training. Rose commented: "It is all personal effort; it depends on each one's competency, experience, knowledge; no standardization or consistency at all". She added that whoever is giving feedback on their assessment tasks is not trained for such a role. She criticised the role of the exam unit because it is not concerned with the content of assessment tools; it only performs an administrative role in the form of receiving, storing, and distributing exams to invigilators and back to graders. Sherry added: "In my context, assessment is ignored and looked over; there is a lack of knowledge among assessors, and the problem is that we do not have a say". Mirjana, Tok, and Rawan had a chance to learn about assessment in previous teaching contexts; however, their preparation was not intensive. Mirjana received informal training in the form of a standardising grading meeting. Rawan took a workshop on designing multiple-choice questions. Tok was lucky enough to be trained on how to base assessment on curriculum and taught material.

#### **4.1.1.3 Teachers' Effort**

The participants mentioned different strategies they adopted to compensate for the lack of assessment literacy development. Mirjana, a critical assessment decision-maker as a reviewer of all APs explained that she depended on her experience as a teacher, colleagues' support, and discussion on different topics, in addition to her instinct about what sounds appropriate. She added: "I also tried to educate myself by searching online for rubrics". Jasmine explained: "We developed our assessment skills based on the little feedback we received, our reflection, previous experience as students, and knowledge of students' culture, abilities, and motivation". Rose tried to develop herself by reading about assessment; however, she admitted that she could not do much due to lack of time. Janset read books on assessment and became a member of testing and evaluation committees. She also checked online courses and trained herself on standardized exams. Talen preferred to learn about assessment through apprenticeship. She practically learnt from her colleagues. Sherry admitted that she had not taken any steps to develop her assessment skills.

#### **4.1.2 Findings Related to Question 1.B**

*Question 1.B How are the assessors involved in assessment tasks in their teaching context?*

Question 1.B focuses on eliciting information about the participants' current involvement in APs and how they were chosen for their assessment tasks. Answer to this question would reveal the assessors' precise roles, and therefore provide clues on the AL level required for such roles. If they were performing crucial tasks in assessment, they would require assessment knowledge equivalent to their high level of indulgence in the assessment process and their selection would have been guided by what they know and can do as assessment designers. Answers to question 1.B was reached after comparing responses provided by the eight participants on two themes related to their assessment involvement in the described context: assessment role and selection criteria.

#### **4.1.2.1 Assessment Role**

Five participants mentioned that they are assigned as team leaders for specific modules (reading, vocabulary, writing, speaking, listening, and grammar). As team leaders, they are required to create summative assessments in the form of mid-term and final exams, in addition to formative assessments in the form of presentations, projects, quizzes, and in-class assignments. These assessment tasks involve setting assessment purpose, selection of assessment material, and designing of assessment items or tasks. The participants also design answer keys and rubrics for grading their designed assessments. They are also required to interpret assessments results and communicate them to different stakeholders (management/learners). Mirjana is head of the program; she is responsible for revising, auditing, approving, and in many cases designing assessments with the team leaders. Sherry and Talen are not team leaders; however, they are involved in designing only formative assessments and assessing learners accordingly.

#### **4.1.2.2 Selection Criteria**

All participants admitted that their selection was not based on specific criteria related to assessment knowledge or skills. Mirjana said that whoever is assigned the role of a team leader should perform the pre-mentioned assessment roles. Jasmine believed she was chosen because she has leadership skills. Tok said that he was approached because the programme management team wanted to change the previous assessment team, and they wanted to add new elements to the team. Rose and Rawan mentioned that their selection was based on availability. Janset guessed that her selection was based on her long experience with teaching. Sherry and Talen believed that the selection is made randomly based on preferences.

#### **4.1.3 Findings Related to Question 2.A**

*Question 2. A. How confident are the assessors with the assessment knowledge base suggested by the TALiP framework?*

Question 2.A focuses on eliciting information about the participants' confidence level with the assessment knowledge base suggested by the TALiP framework. Answers to this question will reveal the teachers'/assessors'

confidence level with the proposed knowledge base, which is a significant fundamental component of their AL. Answer to the above question was reached after comparing responses provided by the eight participants regarding their confidence with the below seven components of assessment knowledge base suggested by the TALiP framework.

#### **4.1.3.1 *Disciplinary & Pedagogical Content Knowledge***

The eight participants expressed their confidence with the five components, which constitute disciplinary and pedagogical knowledge. Most of the participants were generally confident with teaching theories, learning strategies, and learning styles because they were exposed to them as part of their teaching methodology preparation. However, at least five participants did not show high confidence with language theories or language models; they attributed their medium confidence with these two components to lack of training. It was also found that the participants were competent at the components that are directly related to methodology and teaching, which are considered major focus of teacher training. Table 4.1 summarises the participants' responses to and comments on five components associated with disciplinary and pedagogical knowledge.

<b>Sub-Themes of Disciplinary and Pedagogical Content Knowledge</b>	<b>Participants' Confidence with the Sub-Themes</b>
<i>Knowledge of Second Language Learning Theories</i>	Only Tok, Rose and Janset were highly confident with this knowledge due to formal pre-service preparation.
<i>Knowledge of Second Language Teaching Theories</i>	Seven participants were highly confident with this sub-knowledge due to exposure during pre-service or in-service preparation. Mirjana commented: "our education systems primary focus is on teaching methodology". Sherry was the only participant with low confidence due to lack of exposure.
<i>Knowledge of Learning Strategies</i>	Six participants were highly confident with this knowledge because it was part of their methodology training. Only Jasmine and Sherry were medium confident with them.



<i>Knowledge of Learning Strategies</i>	Seven participants were highly confident with this knowledge because they were trained on them. Only Sherry reported being low confident with it.
<i>Knowledge of Language Models</i>	Four participants were highly confident with language models because they learnt about them. The other four participants were not confident due to lack of exposure.

*Table 4.1 The Participants' Responses to and Comments on Five Components Associated with Disciplinary and Pedagogical Knowledge*

#### **4.1.3.2 Knowledge of Assessment Purposes, Content, and Methods**

The eight participants expressed their confidence with the eight theoretical components and sixteen skills, which constitute the knowledge of assessment purpose, content, and methods. Seven participants were not confident with the theoretical components of this knowledge. They showed low confidence with assessment history, philosophies, and impact. The participants showed medium confidence with assessment quality, types, methods, process, and cognition taxonomies. Their medium confidence with these knowledge components is guided by experience and intuitions. Only Tok expressed high confidence with the theoretical components of this knowledge due to formal training. Table 4.2 summarises the participants' responses to and comments on eight components associated with theoretical knowledge of assessment purposes, content and design.

<b>Sub-Themes Associated with Knowledge of Assessment Purposes, Content, and Methods</b>	<b>Participants' Confidence with the Sub-Themes</b>
<i>Knowledge of Language Testing and Assessments History</i>	Six participants reported low confidence with this component because they were never exposed to it; only Tok and Janset were medium confident with it. Tok explained: "I hit on it during the master's program".
<i>Knowledge of Language Testing and Assessment Philosophies:</i>	Seven participants reported low confidence with this knowledge due to lack of exposure. Jasmine commented: "All

	I know is that I assess because we want to make sure that learning took place".
<i>Knowledge of the Impact of Adopting Testing or Assessment Philosophy</i>	Four participants were low confident with this knowledge due to lack of exposure. The other four were medium confident with it guided by experience or their context. Rawan explained "I can only feel how philosophies affect students in the long run".
<i>Knowledge of Major Assessment Qualities</i>	Seven participants showed medium confidence with this knowledge informed by experience, not training. Rawan elaborated: "In my context, they believe it does not matter if assessments show the application of skills as much as they show grades".
<i>Knowledge of Assessment Types</i>	Five participants reported medium confidence with this knowledge. Jasmine and Sherry were low confident with this knowledge. Sherry explained: "I do not know what they involve or how to apply them professionally".
<i>Knowledge of Assessment Methods:</i>	Seven participants were medium confident with this knowledge informed by their teaching experience. Janset explained: "I would love to have someone tell me how to do them professionally".
<i>Knowledge of Test Development Process</i>	Six participants were medium confident with this knowledge. They were familiar with setting test purpose and deciding on the construct, but none of them was aware of how to use test specification. Mirjana explained: "I was not taught these things, but I do some of them while designing assessments". Sherry and Talen were low confident with this knowledge.
<i>Knowledge of Cognition Taxonomies for Item Design</i>	Four participants were confident with this knowledge due to training. Three participants were medium confident with it; they acquired it through experience. Mirjana summarised the situation: "I cannot say I did this because I followed this taxonomy; we are not that professional". Sherry expressed shallow confidence with this knowledge due to ill exposure.

*Table 4.2 The Participants' Responses to and Comments on Eight Components Associated with Theoretical Knowledge of Assessment Purposes, Content, and Methods*

Looking through the participants' responses regarding their confidence with skills associated with the explored knowledge; it is found that they were confident with some skills. They were confident with stating the purpose, writing objectives, aligning assessment objectives with instructions' goals, defining constructs, using alternative methods, providing explicit instructions, providing

examples, and constructing well-laid-out assessments. However, their confidence is guided by their context, as these assessment components are already set for them. Moreover, they reported that they had not received any formal training regarding those skills, and they perform them informed by their teaching experience. As for writing assessment specifications, assessment items, test-format, assuring assessment validity and reliability, they were not confident with these skills. They performed those skills based on intuition and experience with teaching. Table 4.3 summarises the participants' responses to and comments on sixteen skills associated with this knowledge.

<b>Sub-Themes Associated with Practical Skills of Assessment Purposes, Content, and Methods</b>	<b>Participants' Confidence with the Sub-Themes</b>
<i>Stating Assessment Purpose</i>	Seven participants reported confidence with this skill through hands-on teaching. Jasmine and Rawan explained that assessment purposes are set clearly for them by their context; they just need to design assessments that assess them.
<i>Writing Assessment Goals and Objectives</i>	All participants agreed that they are not required to perform this skill because objectives are provided in syllabus and textbooks. They emphasised that they can perform this skill guided by teaching experience, not any formal training.
<i>Aligning Assessment Goals and Objectives with Instructions' Goals and Objectives</i>	Six participants reported confidence with this skill guided by learning outcomes set for them. Jasmine said: "in every assessment we create we have to match the given objectives with the learning outcomes provided by the context".
<i>Defining Language Construct that will be assessed</i>	Seven participants reported confidence with this skill guided by their pre-set contextual outcomes, teaching experience, and familiarity with syllabus and textbooks they are using not by academic training.
<i>Using Alternative Means of Assessments</i>	Eight participants reported confidence with this skill guided by their teaching experience. Mirjana explained: "It is adopted in my teaching context and I also believe in it".
<i>Constructing Test Specifications and Test Forms</i>	Eight participants reported ill-confidence with this skill due to ill exposure. They use a simple form provided by their context.

<i>Writing Selected Response Items</i>	Five participants reported confidence with this skill; while the other three reported medium confidence. Tok & Janset learnt about it through training; the others admitted they perform it intuitively.
<i>Designing Constructed-Response Items</i>	Three participants reported confidence with this skill; however, they have not received any technical input that enables them to validate their designed items. The other five were not high confidence with it.
<i>Constructing Well Laid out Items/Task</i>	Seven participants reported confidence with this skill. None of the participants was given any technical support. Rose explained: "I check how students respond to my tasks to know what is difficult or easy to develop future ones".
<i>Providing Clear Explicit Unambiguous Instructions</i>	Eight participants reported confidence with this skill. None of them received training to perform it. Rose explained: "experience and students' feedback made me learn and develop, so it is much related to my teaching experience."
<i>Providing Examples for Candidates</i>	All participants reported confidence in performing this skill. They do it based on experience.
<i>Including Many Items to Enhance Reliability</i>	Six participants reported confidence with this skill guided by experience. Mirjana explained: "We use many items in the tests; that is how we learnt to do it in our organization". The other two were not confident with this skill.
<i>Writing Test Syllabus to Inform Test Users of Test Formats</i>	All participants were not confident with this skill because they are not required to do so. They just prepare learners in class before assessments through revision classes.
<i>Designing Valid Assessment in Terms of Content and Tasks</i>	The eight participants reported being medium confident with this skill due to ill exposure. Jasmine said: "I do not think these issues are checked in my context; I could not explain if my assessment is valid or not".
<i>Designing Reliable Authentic Fair Practical Ethical and Interactive Assessments</i>	The eight participants reported medium confidence with this skill; they do not have any support to guide them. Sherry acknowledged that she could not guarantee if her assessments have these qualities.
<i>In cooperating Technology in Assessments</i>	Five participants were confident with this skill. They use some programs to help with designing, delivering, or receiving assessments; however, these practices are personal effort initiated by teachers. Three participants

	were not confident with this skill. Janset explained: "our generation was not trained to technology".
--	---

*Table 4.3 The Participants' Responses to and Comments on Sixteen Skills Associated with Knowledge of Assessment Purposes, Content, and Methods*

#### **4.1.3.3 Knowledge of Assessment Grading**

The eight participants expressed their confidence level with six theoretical components and six skills, which constitute the knowledge of assessment grading. Seven participants showed low confidence with all theoretical components of grading knowledge. Only Tok showed some confidence with the proposed grading theoretical knowledge because he was exposed to them during his master's degree. Table 4.4 summarises the participants' responses to the six components associated with theoretical knowledge of grading.

<b>Sub-Themes Associated with Theoretical Knowledge of Assessment Grading</b>	<b>Participants' Confidence with the Sub- Themes</b>
<i>Knowledge of Grading Rational</i>	Five participants were not confident with this knowledge due to ill exposure. Mirjana and Jasmine reported being medium confidence due to little guidance from the institution.
<i>Knowledge of Changing Scores into Meaningful Composites</i>	Five participants reported medium confidence with this knowledge guided by colleagues' support. Sherry reported low confidence due to lack of exposure.
<i>Knowledge of Scaling, Weighting, and Precision of Results</i>	Five participants reported low confidence with this knowledge. Talen commented: "If I know about this, I could stand for my decision because sometimes I feel something is wrong, but I cannot tell why". Mirjana and Jasmine reported medium confidence as a result of having a hand on grading.
<i>Knowledge of Scoring Techniques for Objectively-Marked Testing</i>	Four participants reported medium confidence with this knowledge based on intuition. Sherry and Rose expressed ill confidence with it.
<i>Knowledge of the Nature, Purpose, and Design of Scoring Rubrics for Subjectively Marked Tests</i>	Five participants were confident with this knowledge; however, they design scoring rubrics based on their intuitions and feedback from colleagues. Sherry and Talen expressed ill confidence with this knowledge.

<i>Knowledge of Consistency and Moderation &amp; Principles for Judgment-Based Assessments:</i>	Seven participants expressed low confidence with this knowledge. Rawan explained: "I am not involved in this, and I do not know how to do it".
---	--

*Table 4.4 The Participants' Responses to and Comments on Six Components Associated with Theoretical Knowledge of Assessment Grading*

It is also found that the participants are not confident with most grading skills specifically those related to statistical issues and quantification of data: calculating consistency of scores, quantifying performance, calculating reliability or scorer's reliability. Despite reporting being skilled in designing rubrics for open-ended responses and close-ended responses; they also reported that their practice is not based on any knowledge, and they only gained it through practice. Moreover, their requirement for a professional body to validate what they do contradicts with the confidence level some of them reported with these skills. Table 4.5 summarises the participants' responses to and comments on six skills associated with this knowledge.

<b>Sub-Themes Associated with Skills of Assessment Grading</b>	<b>Participants' Confidence with the Sub- Themes</b>
<i>Constructing Scoring Schemes that Quantify Learners' Performance</i>	Five participants reported medium confidence with this skill. Janset explained: "I cannot tell or stand for my point, and I do not believe what I do is scientific". Sherry, Rose, and Talen reported ill confidence with it due to lack of training.
<i>Designing Scoring Keys with Explicit Instructions for Closed-Ended Responses</i>	Seven participants reported confidence with this skill; they developed it on the job. Only Sherry admitted being low confident with this skill; she justified: "I am not sure if what I do is right or wrong".
<i>Designing Scoring Rubrics with Explicit Instructions for Open-Ended Responses</i>	Seven participants reported confident with this skill; however, they were never trained to do it. Rawan explained: "I got it by experience and self-learning". Sherry said that she is low confident with it and she is not even familiar with its terms.
<i>Calculating Consistency of Scores/ Standard Error / Scoring Reliability for both Closed- and Open-Ended Responses:</i>	The eight participants reported low confidence with this skill. Rawan said it is not needed in their context; while Tok said that he would like to read more about it.

<i>Calculating Scorer Reliability</i>	The eight participants reported low confident with this skill due to lack of exposure and unfamiliarity with the topic.
<i>Designing Training Workshops for Scorers</i>	The eight participants reported low confidence with this skill. Jasmine, Sherry, and Talen argued that they need this kind of training themselves before thinking of doing it to others.

*Table 4.5 The Participants' Responses to and Comments on Six Skills Associated with Knowledge of Assessment Grading*

#### **4.1.3.4 Knowledge of Assessment Feedback**

The eight participants expressed their confidence level with three theoretical components and eleven skills, which constitute the knowledge of providing assessment feedback. The participants were confident with general principles of feedback, which they acquired through experience, not formal training; however, they lacked the mechanism of providing feedback to learners in the form of types or forms. Table 4.6 summarises the participants' responses to and comments on the three components associated with theoretical knowledge of assessment feedback.

<b>Sub-Themes Associated with Theoretical Knowledge of Assessment Feedback</b>	<b>Participants' Confidence with the Sub-Themes</b>
<i>Knowledge of Purposes and Principles of Feedback</i>	All participants were confident with this knowledge guided by their teaching experience. They acknowledged its power in developing learning.
<i>Knowledge of Different Types of Feedback and Their Functions</i>	Four participants were not confident with this knowledge. Rawan explained: "we do not have the time to customize or personalize feedback". Sherry, Rose, and Talen were not familiar with the types. Tok was the only participant who showed confidence with it.
<i>Knowledge of Strengths and Weaknesses of Kinds of Feedback</i>	All participants did not show confidence with this knowledge. They mentioned that they are not familiar with it.

*Table 4.6 The Participants' Responses to and Comments on Three Components Associated with Theoretical Knowledge of Assessment Feedback*

The participants showed confidence with feedback skills, which are associated with teaching, such as setting goals for learning tasks, constructing learning tasks, sharing criteria for success, showing learning gaps, providing motivating feedback, recording learning development, and evaluating learners' success. These skills are related to teaching objectives and learning outcomes, which are set clearly by their teaching context. Moreover, most participants were not confident with skills, such as intervening techniques, scaffolding exercises, using internet resources for feedback, using various tools to monitor learning, and recording development techniques because these skills require technical training. Table 4.7 summarises the participants' responses to and comments on eleven skills associated with this knowledge.

<b>Sub-Themes Associated with Skills of Assessment Feedback</b>	<b>Participants' Confidence with the Sub- Themes</b>
<i>Setting Goals and Constructing Tasks to move Learners towards Them</i>	Seven participants reported being confident with this skill guided by their teaching experience and context, which set them clearly for them. Sherry reported inability to set educational goals, but she can construct tasks that enable learners to achieve them.
<i>Communicating to Learners the Criteria for Success through Different Modes</i>	Seven participants reported being confident with this skill informed by their teaching experience. Jasmine explained: "we usually do that at the beginning of the semester and before each assessment".
<i>Showing Learners, the Gap between their Current Stage and the Next/Final Stage</i>	Six participants were confident with this skill, although they are not required to do so, they do it intuitively. Sherry and Janset reported being confident with the concept itself, but they are not confident with techniques to apply this skill.
<i>Performing "Interventionist"/ "Interactionist" Role</i>	Two participants reported confidence with this skill informed by their teaching experience. Three participants mentioned that they are familiar with the concepts, but they lack precise techniques to apply the skill. Sherry and Talen reported unfamiliarity with this skill.
<i>Scaffolding Learning through Mediating Exercises</i>	Six participants reported being unfamiliar with these skills or practices. Jasmine said: "I am not aware of how the task is designed". Only Rawan and Tok mentioned that they have some background, but they have never used them.



<i>Using Internet Resources to Address Students' Needs</i>	Six participants reported medium confidence with this skill. Mirjana and Sherry mentioned that they never used this skill as a mean for providing formative feedback.
<i>Using Assessment Methods to Monitor learning</i>	Six participants reported medium confidence with this skill guided by their teaching experience. Rose said that she does it as a self-initiated practice; while, Sherry admitted that she had never tried it.
<i>Providing Encouraging Feedback to Modify the Learning Process</i>	All participants reported confidence with this skill guided by their teaching experience. Sherry explained: "I do not have the theoretical preparation, but I do this out of my ethical perspective as a teacher'.
<i>Recording and Reporting Student Development</i>	Two participants reported being confident with this skill, and four reported being medium confidence with it. Mirjana and Sherry said that they have never tried this. Rawan reported just using learners' scores on the grade book to monitor their development.
<i>Evaluating Learners' Success in Achieving Learning Goals</i>	Seven participants reported being confident with this skill; however, they reported using only learners' scores to evaluate their achievement. Rose argued: "the best way to perform this skill is to keep portfolios; however, we are not trained to do so".
<i>Improving Instruction based on Assessment Results/Feedback</i>	Seven participants reported being confidence with this skill influenced by their teaching experience. Jasmine explained: "sometimes based on the results we develop instructions". Only Sherry reported lack of this skill.

*Table 4.7 The Participants' Responses to and Comments on Eleven Skills Associated with Knowledge of Assessment Feedback*

#### **4.1.3.5 Knowledge of Peers- and Self-Assessments**

The eight participants expressed their confidence level with two theoretical components and five skills, which constitute the knowledge of peer- and self-assessment. Four participants were not confident with benefits of peer-and self-assessments; the other half were aware of its benefits; however, this awareness was not based on academic input; it was built up during teaching experience. Almost all participants were not confident with peer- and self-assessment strategies; they did not know its techniques or how to apply them. One participant

was against peer and self-assessment principles and techniques; she did not believe in their power in enhancing learning. Table 4.8 summarises the participants' responses to and comments on the two components associated with the theoretical knowledge of peer and self-assessment.

<b>Sub-Themes Associated with Theoretical Knowledge of Peers- and Self-Assessments</b>	<b>Participants' Confidence with the Sub- Themes</b>
<i>Knowledge of Benefits of Involving Students in Assessment</i>	Four participants were confident with this knowledge informed by their teaching experience. The other participants were not confident with it. Talen considered it unreliable knowledge.
<i>Knowledge of Self- and Peer Assessment Strategies</i>	Five participants were medium confident with this knowledge. Jasmine explained: "I know about it, but I do not know how to apply it". Sherry, Rose, and Talen were not confident with these strategies. Talen justified: "I cannot trust students; teachers should control assessments".

*Table 4.8 The Participants' Responses to and Comments on Two Components Associated with Theoretical Knowledge of Peers- and Self-Assessments*

Almost all participants were not confident with skills related to self-and peer-assessments, such as training learners to participate in assessments, training learners to create their assessment criteria to assess themselves or peers, or creating portfolios to keep a record of learners' assessments. They showed confidence with other sub-skills, such as involving learners in noticing their problems and their peers' problems, in addition to providing them with criteria or rubrics to assess themselves or peers. However, it is noticed that their confidence with these skills was informed by their teaching experience, not a technical input. Table 4.9 summarises the participants' responses to and comments on five skills associated with this knowledge.

<b>Sub-Themes Associated with Skills of Peers- and Self-Assessments</b>	<b>Participants' Confidence with the Sub- Themes</b>
<i>Training Learners to Participate in Assessment</i>	Three participants reported medium confidence with this skill. The others were low confidence with it. Rose argued that she does not have time, while

	Talen said that she does not believe in it. Only Tok mentioned that he is confident with it.
<i>Using Self- and Peer-assessments to Promote Learners' Ability to Notice their Learning Gaps</i>	Seven participants reported confidence with this skill informed by their teaching practices. Jasmine explained: "I want them to notice their problems away from exams stress". Only Talen reported low confidence with this skill.
<i>Providing Learners with Criteria to Judge their Work and their Peers' Work</i>	Seven participants reported confidence with this skill. They provide learners with grading rubrics to grade their work or their colleagues' work. Jasmine justified: "I feel the more guided the task is, the more productive the outcome will be". Only Sherry reported ill confidence with it.
<i>Encouraging Learners to Produce their Rating Criteria</i>	All participants reported ill confidence with this skill. They have not tried it. Jasmine commented: "I do not believe they are at that competent level; I cannot trust it fully".
<i>Training Learners to use Portfolios, Diaries, Continuous Cards, Audios/Video Diaries to Monitor their Learning and Keep its Records</i>	Only Tok and Rawan reported confidence with this skill. The other six participants were not confident with it. Jasmine explained: "there should be certain circumstances for it to work: teachers should be willing, classrooms should be small, and the purpose of the assessment should be different."

*Table 4.9 The Participants' Responses to and Comments on Five Skills Associated with Knowledge of Peers- and Self-Assessments*

#### **4.1.3.6 Knowledge of Assessment Interpretation and Communication**

The eight participants expressed their confidence level with two theoretical components and ten skills, which constitute the knowledge of assessment interpretation and communication. Most participants were low confident with the theoretical concepts of assessment interpretation and communication. They have not received any training, and they were not requested to practice them in their current context. They perform fundamental interpretation informed by self-initiation and based on teaching experience. Table 4.10 summarises the participants' responses to and comments on two associated with the theoretical knowledge of assessment interpretation and communication.

<b>Sub-Themes Associated with Theoretical Knowledge of Assessment Interpretation and Communication</b>	<b>Participants' Confidence with the Sub Themes</b>
<i>Knowledge of Assessment Reliability (Dependability/ Item Analysis/Reliability Threat)</i>	Six participants reported low confidence with this knowledge due to lack of exposure. Only Mirjana and Tok showed some confidence with item analysis but they were not familiar with reliability threats and dependability.
<i>Knowledge of Assessment Validity</i>	Four participants were not familiar with this concept. The other four participants reported medium confidence with some types of validity. Rose said: "I know some of them, but I do not know the terminology".

*Table 4.10 The Participants' Responses to and Comments on Two Components Associated with Theoretical Knowledge of Assessment Interpretation and Communication*

It was also found that the participants were low confident with all skills related to assessment interpretation and any skill that requires calculation and statistical procedures, such as item difficulty, discrimination, reliability, and validity. They were not trained to perform these tasks, and they were not required to do so in their teaching context. The participants showed medium confidence with skills related to results communication, such as articulating and communicating assessment results to different stakeholders and participating in communities regarding assessment reforms, or assisting learners with interpreting their assessment results; however, they perform these skills informed by their teaching experience. Table 4.11 summarises the participants' responses to and comments on ten skills associated with this knowledge.

<b>Sub-Themes Associated with Skills of Assessment Interpretation and Communication</b>	<b>Participants' Confidence with the Sub-Themes</b>
<i>Calculating Item Difficulty /Item Discrimination for Close-Ended Items</i>	Five participants reported low confidence with this skill, while the other three participants reported being medium confident with it. Mirjana explained: "I do them in a very basic manner".
<i>Calculating Test Reliability and Inter/Intra-Rater Reliability</i>	Five participants reported low confidence with this skill. Three participants mentioned they know about them, but they cannot do the calculations.

<i>Calculating Means/Modes/Medians/ Bell curves in Large-Scale Tests for Data Interpretation</i>	Eight participants reported low confidence with this knowledge due to lack of training. Even Tok said he does not know how to use them.
<i>Investigating Validity Using Statistical Procedures</i>	Six participants reported low confidence with this skill. Two participants know the concepts, but they cannot do any calculations. Rawan commented: "when it comes to calculation and static I need help".
<i>Inferring Students' Strength and Weakness Based on Collected Data</i>	Six participants reported medium confidence with this knowledge, and two reported low confidence with it. They developed it through teaching and daily contact with students. Janset commented: "I do them intuitively not in terms of statistic at all".
<i>Articulating Interpretation of Norm- and Criterion-Referenced Assessment Results</i>	The eight participants reported medium confidence with this skill guided by their teaching experience and knowledge of their students. They can only interpret results from criterion-referenced assessments, but not from norm-referenced.
<i>Communicating Interpretations using Different Reporting Tools</i>	Seven participants reported competence with communicating results to different stakeholders based on their teaching experience and criteria required by their teaching context. They communicate grades, but they do not conduct meeting or counselling sessions.
<i>Using Software (Statistical Package) to Communicate Results</i>	The eight participants reported low confidence with this skill. Mirjana explained "we only do basic manual stuff".
<i>Participating in Committee or School-Wide Discussions about Reforms in Assessment Related Issues</i>	Six participants reported confidence with this skill, and two reported medium confidence. The participants showed interest in taking part in this practice because they want their voice to be heard. Jasmine explained: "we review exams and reflect on them, but not on the higher management level".
<i>Coaching Students to Analyse Assessment Results, Track their Learning, and Plan Next Steps</i>	Five participants reported medium competence with this skill, and three participants reported low confidence with it. They were not required to perform it. Jasmine explained: "I do it randomly, but I have not applied it seriously."

*Table 4.11 The Participants' Responses to and Comments on Ten Skills Associated with Knowledge of Assessment Interpretation and Communication*

#### **4.1.3.7 Knowledge of Assessment Ethics**

The eight participants expressed their confidence level with four theoretical components and sixteen skills, which constitute the knowledge of assessment ethics. All participants were not confident with the theoretical

knowledge related to assessment ethics. Their knowledge is informed by teaching experience and what sounds appropriate to them. They were not exposed to any academic input that would validate their current knowledge. The only knowledge they showed confidence with is protecting the confidentiality and integrity of assessment, which they acquired on the job. Table 4.12 summarises the participants' responses to and comments on four components associated with the theoretical knowledge of assessment ethics.

<b>Sub-Themes Associated with the Theoretical Knowledge of Assessment Ethics</b>	<b>Participants' Confidence with the Sub-Themes</b>
<i>Knowledge Critical Language Testing: Power, Ethics, and Fairness</i>	Five participants reported low confidence with this knowledge. Jasmine commented: "For me, ethics in assessment is related to plagiarism, cheating, originality, fair grading". Three participants reported medium confidence with these concepts.
<i>Knowledge of Equity/ Non-Discrimination/ Inclusion/ Social Justice Principles</i>	Four participants reported low confident with this knowledge. Janset commentated: "I believe in these things, but I do not know how to work towards them". The other four participants reported medium confident with it; they deal with it intuitively.
<i>Knowledge of Codes/ Concepts of Professional Morality</i>	Four participants reported low confident with this knowledge. They deal with it intuitively; the other participants reported medium confidence; they acquired it through practice.
<i>Knowledge of Legal/ Ethical responsibilities Concerning Use/ Storage/ Dissemination of Results</i>	Almost all participants reported confidence with this knowledge because they learnt it on the job. Rose commented: "that is what organizations focus on and nothing else".

*Table 4.12 The Participants' Responses to and Comments on Four Components Associated with Theoretical Knowledge of Assessment Ethics*

It was also found that they were confident with some sub-skills of ethics knowledge, such as following context-adopted ethics guideline, treating learners with respect, adopting fair, non-discriminatory practices, avoiding controversial, inflammatory material, informing learners about assessment features, involving colleagues in decision making, being open to constructive feedback, and using multi-methods of assessments. However, their confidence with these skills is

informed by their intuition and experience with teaching; they were not formally guided or academically trained to perform these skills.

On the other hand, they reported medium and low confidence with other skills of assessment ethics knowledge, such as the ability to develop assessment through writing clear purposes, setting criterion, designing assessment specification, providing evidence of assessment fairness, calculating validity and reliability, inferring decisions from assessment results, using results to develop learning through feedback, evaluating assessment washback, evaluating consequential validity, involving learners in assessment practices, and critiquing the power of standardized assessments. The participants' ill-confidence with these skills is attributed to lack of knowledge, training, formal academic guidance. They perform some of these skills if required by their context informed by their intuition and experience with teaching. Table 4.13 summarises the participants' responses to and comments on sixteen practical components associated with this knowledge.

<b>Sub-Themes Associated with Skills of Assessment Ethics</b>	<b>Participants' Confidence with the Sub- Themes</b>
<i>Observing Ethics in Assessment Practices</i>	The eight participants mentioned that they are confident with this skill; they abide by their context adopted ethical guidelines. Janset explained: "I do what is required and given to me as per provided checklists."
<i>Treating all Users of Assessment with Respect</i>	The eight participants reported being confident with this skill. Rose commented: "This is done intuitively; we are educators".
<i>Providing Fair/Non-Discriminatory practices and Assuring Distributive Justice</i>	Six participants mentioned that they are confident with this skill based on their experience as teachers. Janset and Rawan were not confident with this skill. They mentioned they need to validate their practices because it is not based on technical input.
<i>Avoiding Construct-Irrelevant/ Controversial/ Inflammatory/ Offensive/ Upsetting Material</i>	Almost all participants reported confidence with this skill. They do it intuitively guided by their teaching experience.
<i>Informing Learners of What, How, and Why of Assessments</i>	The eight participants reported confidence with this skill. They would practice it intuitively even if their

	context did not require it. Rawan commented: "I like to be transparent and I enjoy doing it".
<i>Involving other Colleagues in Assessment Decision/ Practices</i>	The eight participants reported confidence with this skill even though they are not encouraged to involve other teachers in assessment decisions.
<i>Being Open to Constructive Feedback from Colleagues</i>	The eight participants reported confidence with this skill; they all believed in the benefit of receiving feedback from their colleagues for the sake of development.
<i>Using a Range of Assessment Approaches</i>	The eight participants reported confidence with this skill guided by their teaching experience and knowledge of their learners' preferences, abilities and needs.
<i>Defining a Clear Purpose/ Developing Specifications/ Evaluating Content and Conducting Field Test Examination</i>	The eight participants did not report high confidence with this skill. They do it informed by their teaching experience; however, they lack practical training that enables them to go through assessment design process competently or at least defend their constructed assessment.
<i>Providing Evidence of Fairness via Statistical Procedures such as Validity and Reliability</i>	The eight participants reported low confidence with this skill. They explained that they do not practice it in their context because these concepts are not paid attention to or checked by supervisors.
<i>Informing the Inferences and Decisions that Derive from Assessments Scores</i>	The eight participants reported medium confidence with this skill guided by their teaching experience and knowledge of their learners. Rawan commented: "I can form my interpretation but not from a statistic or theoretical-based knowledge".
<i>Using Assessment Results for Feedback to Influence Learning</i>	Almost all participants reported medium confidence with this skill; they reported doing it without being asked. Jasmine said that she always questions her teaching and assessment practice.
<i>Evaluating Assessments Washback on Learning, Teaching, Curricula, and Institutions</i>	All participants reported medium confidence with this skill. They explained that they are not involved in this process. Sherry said, "we cannot evaluate; we are not that competent, we can judge".
<i>Judging the Consequences of Assessment in own Context through Examining "Consequential Validity."</i>	All participants reported low confidence with this skill. Ill confidence is due to ill-training. Rawan added that this requires "coordination" among various departments, which they lack in their context.



<i>Allowing Students to Share their Voices about Assessment:</i>	Almost all participants reported medium confidence with this skill because they do not have sufficient knowledge to perform it professionally. Talen was totally against this skill; she argued: "This is my job, from where are they going to get this knowledge?"
<i>Critiquing the Impact and Power Standardized Tests</i>	All participants reported low confidence with this skill due to lack of experience and training.

*Table 4.13 The Participants' Responses to and Comments on Sixteen Skills Associated with the Knowledge of Assessment Ethics*

#### **4.1.4 Findings Related to Question 2.B**

*Question 2.B What are the assessors' views regarding their need for the different components of assessment knowledge base suggested by the TALiP framework?*

The fourth research question focuses on eliciting information about the participants' beliefs about what is needed from the proposed knowledge base. This step is crucial because they will justify reasons for their views. After all, not all assessment knowledge is applicable or required for all contexts. Answers to this question will filter the proposed knowledge base of AL as per the assessors' conception of what is needed or not needed for their current assessment tasks, which will, in turn, reveal another advanced level of their AL as per the TALiP model. Answer to the above question was reached after comparing responses provided by the eight participants regarding their need for the below seven components of assessment base knowledge suggested by the TALiP framework.

##### **4.1.4.1 Need for Disciplinary Knowledge & Pedagogical Content Knowledge**

The participants expressed their need for the five components, which constitute disciplinary and pedagogical knowledge and specified in table 4.14.

<i>Knowledge of Second Language Learning Theories</i>	<i>Knowledge of Second Language Teaching Theories</i>	<i>Knowledge of Learning Strategies</i>	<i>Knowledge of Learning Styles</i>	<i>Knowledge of Language Models</i>
---	---	---	-------------------------------------	-------------------------------------

*Table 4.14 Components of Disciplinary and Pedagogical Knowledge*

All participants expressed interest in these knowledge components because they believe they are integrated with and inseparable from their assessment knowledge. They also believe that disciplinary and pedagogical knowledge guide their assessments decisions related to the selection of tools, items, and content and therefore competency at this knowledge would affect assessment practices positively. Rose commented: "if I change the adopted learning theory, the assessment will change accordingly". Mirjana explained: "if I adopt a communicative approach, I should focus on fluency not on the accuracy or grammatical errors". All participants expressed their need for knowledge of learning strategies and styles. Mirjana summarised their responses: "we need to give them more individualized chances to come up with fair assessments". There was a consensus among the eight participants regarding their need for knowledge of language models. They explained that these are their teaching tools and the content of their assessments.

#### **4.1.4.2      *Need for Knowledge of Assessment Purposes, Content and Methods***

The participants expressed their need for the eight theoretical components and sixteen skills, which constitute the knowledge of assessment purpose, content and methods. They first expressed their need for the below theoretical knowledge specified in table 4.15.

<i>Knowledge of Language Testing and Assessments History</i>
<i>Knowledge of Language Testing and Assessment Philosophies</i>
<i>Knowledge of the Impact of Adopting Testing or Assessment Philosophy</i>
<i>Knowledge of Major Assessment Qualities</i>
<i>Knowledge of Assessment Types</i>
<i>Knowledge of Assessment Methods</i>
<i>Knowledge of Test Development Process</i>
<i>Knowledge of Cognition Taxonomies for Item Design</i>

*Table 4.15 Theoretical Components of Assessment Purpose, Content and Methods*

Five participants reported that they need to know about assessment history to benefit from past expertise. However, Mirjana, Tok, and Rose believed they do not need to know it. Mirjana commented: "we need to focus on the current situation because assessment theories change". They all supported knowledge of assessment philosophies and their impact on assessments because it will enable them to defend their own beliefs against imposed ones and to protect assessment practices from manipulation. They all considered knowledge of assessment qualities, types, purpose, design, methods, and item writing core ones to perform their assessment tasks. Jasmine explained: "we need to keep assessment qualities as a checklist because our designed items will change accordingly". All participants acknowledged the need for knowledge about assessment types, specifically standardized assessments, even if they are not currently involved in them. Mirjana explained: "some of our students are involved in such assessments; we need to guide them". Rawan added: "they are our competitors; we need to critique and be able to form our own personalized, individualized knowledge on how thing should be assessed". There was a consensus on the need for assessment methods. Jasmine explained: "we cannot depend only on tests, but using a mixture can tell a lot if you are after the idea of addressing learners' differences, needs, and preferences". All participants believed they need to know about the assessment design process. Jasmine explained: "I believe this is the handbook for an assessor". Mirjana commented: "if I were trained about such things, I would have been more confident about them".

It was also found that all participants expressed their need for all skills of this knowledge listed in table 4.16; however, they required guidance on how to perform them to assure fairness and ethicality of their assessment and to avoid negative washback. Janset explained: "these skills standardize our work; when things are well-settled, I have more confidence and trust in the validity of my assessment, and I can defend them". They also required technical training specifically on test specification, item writing, reliability, and validity to justify all their assessment decisions and be able to create "test forms" that originate from the same specification to secure equivalency, reliability and validity. Janset argued: "I need practical learning; it will complete the picture and make me

satisfied with what I do". Jasmine added: "we need an expert to validate our practices; this is crucial for our assessment fairness and ethicality".

<i>Stating Assessment Purpose</i>
<i>Writing Assessment Goals and Objectives</i>
<i>Aligning Assessment Goals and Objectives with Instructions' Goals and Objectives</i>
<i>Defining Language Construct that will be Assessed</i>
<i>Using Alternative Means of Assessments</i>
<i>Constructing Test Specifications and Test Forms</i>
<i>Writing Selected Response Items</i>
<i>Designing Constructed-Response Items</i>
<i>Constructing Well Laid out Items/Task</i>
<i>Providing Clear Explicit Unambiguous Instructions</i>
<i>Providing Examples for Candidates</i>
<i>Including Many Items to Enhance Reliability</i>
<i>Writing Test Syllabus to Inform Test Users of Test Formats</i>
<i>Designing Valid Assessment in Terms of Content and Tasks</i>
<i>Designing Reliable Authentic Fair Practical Ethical and Interactive Assessments</i>
<i>In cooperating Technology in Assessments</i>

Table 4.16 Skills Associated with Assessment Purpose, Content, and Methods

#### **4.1.4.3 Need for Knowledge of Assessment Grading**

The eight participants expressed their need for the six theoretical components and six skills, which constitute the knowledge of assessment grading listed in table 4.17 because they are involved in grading tasks, and they cannot perform it adequately without this knowledge. Janset argued: "we put effort on assessments, but to lose control on it due to ill knowledge of scoring is shameful". It was also found that all participants consider mastering grading skills essential to defend their grading practices and be fair to learners. Rose commented: "I do not know them; I need to know to be able to implement them".

<i>Knowledge of Grading Rational</i>
<i>Knowledge of Changing Scores into Meaningful Composites</i>
<i>Knowledge of Scaling, Weighting, and Precision of Results</i>
<i>Knowledge of Scoring Techniques for Objectively-Marked Testing</i>
<i>Knowledge of the Nature, Purpose, and Design of Scoring Rubrics for Subjectively Marked Tests</i>
<i>Knowledge of Consistency and Moderation &amp; Principles for Judgment-Based Assessments</i>
<i>Constructing Scoring Schemes that Quantify Learners' Performance</i>
<i>Designing Scoring Keys with Explicit Instructions for Closed-Ended Responses</i>
<i>Designing Scoring Rubrics with Explicit Instructions for Open-Ended Responses</i>
<i>Calculating Consistency of Scores/ Standard Error / Scoring Reliability for both Closed- and Open-Ended Responses</i>
<i>Calculating Scorer Reliability</i>
<i>Designing Training Workshops for Scorers</i>

*Table 4.17 Components of Assessment Grading Knowledge*

#### **4.1.4.4 Need for Knowledge of Providing Formative Feedback**

The eight participants expressed their need for the three theoretical components and eleven skills, which constitute the knowledge of providing assessment feedback listed in table 4.18. All teachers acknowledged their need for feedback knowledge to provide constructive feedback on learners' work. Janset explained that she could not give the same feedback to all learners. She commented: "I cannot put everybody in the same basket; unfortunately, most of the time, we focus only on evaluation". Rose added: "I need to know about them if we want to do things in the right way". There was a consensus among all participants regarding their need for the proposed feedback skills. They believed they need these skills because they enhance learners' accomplishment and provides a clear image of everything. They added that if learners are left alone without guidance, they will not improve; they will just pay attention to the results of assessments. Rawan commented: "I do not teach for exams. I teach for life. Scores should not be the end mean of assessments". Janset added: "assessment is not only for learners; it is also for teachers, management, and the whole system".

<i>Knowledge of Purposes and Principles of Feedback</i>
<i>Knowledge of Different Types of Feedback and Their Functions</i>
<i>Knowledge of Strengths and Weaknesses of Kinds of Feedback</i>

<i>Setting Goals and Constructing Tasks to move Learners towards Them</i>
<i>Communicating to Learners the Criteria for Success through Different Modes</i>
<i>Showing the Learners the Gap between their Current Stage and the Next/Final Stage</i>
<i>Performing "Interventionist"/ "Interactionist" Role</i>
<i>Scaffolding Learning through Mediating Exercises</i>
<i>Using Internet Resources to Address Students' Needs</i>
<i>Using Assessment Methods to Monitor learning</i>
<i>Providing Encouraging Feedback to Modify the Learning Process</i>
<i>Recording and Reporting Student Development</i>
<i>Evaluating Learners' Success in Achieving Learning Goals</i>
<i>Improving Instruction based on Assessment Results/Feedback</i>

Table 4.18 Components of Assessment Feedback Knowledge

#### **4.1.4.5 Need for Knowledge of Peer- and Self-Assessments**

The participants expressed their need for the two theoretical components and five skills, which constitute the knowledge of peer- and self-assessment listed in table 4.19. Seven participants expressed their need for the theoretical components of this knowledge. They believe that acquiring this knowledge will make them empowers their learners and train them to have a hand on their learning. Besides, they explained that implementing these principles in their contexts can relieve stress and shift the focus to learners. Janet explained: "it is also motivating; it makes them responsible". Rawan added: "it could be an alternative for face to face feedback to save time and manage classes with a high number of students". Only Talen did not believe in involving learners in assessment, as she argues assessment is a critical role that should be taken care only by teachers. Talen said that she would never use it because she cannot trust learners with it. She commented: "students use it to go out of the teaching mood". The same seven participants supported the need for all proposed skills except for encouraging learners to create their rubrics because they believe learners do not possess the required knowledge or skill to do so. Talen argued: "they are not up to this level; from where would they get this knowledge?"

<i>Knowledge of Benefits of Involving Students in Assessment</i>
<i>Knowledge of Self- and Peer Assessment Strategies</i>
<i>Training Learners to Participate in Assessment</i>
<i>Using Self- and Peer-assessments to Promote Learners' Ability to Notice their Learning Gaps</i>
<i>Providing Learners with Criteria to Judge their Work and their Peers' Work</i>

<i>Encouraging Learners to Produce their Rating Criteria</i>
<i>Training Learners to use Portfolios, Diaries, Continuous Cards, Audios/Video Diaries to Monitor their Learning and Keep its Records</i>

*Table 4.19 Components of Peer- and Self-Assessments Knowledge*

#### **4.1.4.6 Need for Knowledge of Assessment Interpretation and Communication**

The participants expressed their need for the two theoretical components and ten skills, which constitute the knowledge of assessment interpretation and communication listed in table 4.20. Almost all participants expressed the need for this knowledge to validate and defend their practices and to be able to present fair, reliable assessments. Rawan said: "I need to make sure that whatever grades I give reflect their knowledge". Talen argued: "I feel something is wrong, but I do not know how to critique or defend my points". Seven participants expressed their need for skills related to this knowledge because all stakeholders have the right to understand assessment results and for learning development; however, they requested more training on how to perform it, as they believe it is essential for learners' improvement and for giving them a voice in assessment decisions. Rose added that this practice is related to transparency because teachers have to provide precise data and interpretations. Janset commented: "This knowledge helps us to critique our assessments for improvements". Mirjana did not believe in learning about those skills because she believes other bodies need to be involved in it. She argued that classroom assessment they are involved in does not require those skills. Regarding result communication, they all supported this skill.

<i>Knowledge of Assessment Reliability (Dependability/ Item Analysis/Reliability Threat):</i>
<i>Knowledge of Assessment Validity</i>
<i>Calculating Item Difficulty /Item Discrimination for Close-Ended Items</i>
<i>Calculating Test Reliability and Inter/Intra-Rater Reliability</i>
<i>Calculating Means/Modes/Medians/ Bell Curves in Large-Scale Tests for Data Interpretation</i>
<i>Investigating Validity Using Statistical Procedures</i>
<i>Inferring Students' Strength and Weakness Based on Collected Data</i>
<i>Articulating Interpretation of Norm- and Criterion-Referenced Assessment Results</i>
<i>Communicating Interpretations using Different Reporting Tools</i>

<i>Using Software (Statistical Package) to Communicate Results</i>
<i>Participating in Committee or School-Wide Discussions about Reforms in Assessment Related Issues</i>
<i>Coaching Students to Analyse Assessment Results, Track their Learning, and Plan Next Steps</i>

Table 4.20 Component of Assessment Interpretation and Communication Knowledge

#### 4.1.4.7 **Need for Knowledge of Assessment Ethics**

The participants expressed their need for the four theoretical components and sixteen skills, which constitute the knowledge of assessment ethics listed in table 4.21. They all expressed the need for this knowledge, as it would assist them in implementing ethics principles in their assessments and consequently producing fair assessments. Janset explained that this knowledge would empower assessors and give them the chance to utter their beliefs and rights. Rose added: "implementing these principles will encourage learners to have a say in their assessment". Only one participant believed that such knowledge should not be taught; it could be acquired by experience. All participants expressed their needs for all proposed skills. They find them essential for their assessments practices. However, they emphasised their need for training, which can support their practices and validate them.

<i>Knowledge Critical Language Testing: Power, Ethics, and Fairness</i>
<i>Knowledge of Equity/ Non-Discrimination/ Inclusion/ Social Justice Principles</i>
<i>Knowledge of Codes/ Concepts of Professional Morality</i>
<i>Knowledge of Legal/ Ethical responsibilities Concerning Use/ Storage/ Dissemination of Results</i>
<i>Observing Ethics in Assessment Practices</i>
<i>Treating all Users of Assessment with Respect</i>
<i>Providing Fair/ Non-Discriminatory Practices and Assuring Distributive Justice</i>
<i>Avoiding Construct-Irrelevant/ Controversial/ Inflammatory/ Offensive/ Upsetting Material</i>
<i>Informing Learners of What, How, and Why of Assessments</i>
<i>Involving Other Colleagues in Assessment Decision/ Practices</i>
<i>Being Open to Constructive Feedback from Colleagues</i>
<i>Using a Range of Assessment Approaches</i>
<i>Defining a Clear Purpose/ Developing Specifications/ Evaluating Content and Conducting Field Test Examination</i>
<i>Providing Evidence of Fairness via Statistical Procedures such as Validity and Reliability</i>
<i>Informing the Inferences and Decisions that Derive from Assessments Scores</i>



<i>Using Assessment Results for Feedback to Influence Learning</i>
<i>Evaluating Assessments Washback on Learning, Teaching, Curricula, and Institutions</i>
<i>Judging the Consequences of Assessment in own Context through Examining "Consequential Validity"</i>
<i>Giving Students the Opportunity to Share their Voices about Assessment</i>
<i>Critiquing the Impact and Power Standardized Tests</i>

*Table 4.21 Components of Assessment Ethics Knowledge*

#### **4.1.5 Findings Related to Question 3.A**

##### ***Question 3.A How are assessments practised in the participants' teaching context?***

Question 3A focuses on eliciting information about the participants' context-adopted assessments features. The answer to the question will help in exploring and understanding a third level of the participants' AL, which is their understanding of their context adopted APs. Answers to the above question were reached after comparing responses provided by the eight participants on thirteen themes that resulted from the data analysis related to this question.

##### **4.1.5.1 Assessment Purpose**

The participants reported that the assessment purpose in their teaching context is checking the achievement of pre-set learning goals and evaluating their performance to infer if they can pass the program or not. Ongoing summative assessments evaluate the achievements of these goals and provide grades. In the end, an accumulative grade is calculated to assist in making the decision.

##### **4.1.5.2 Assessment Source**

The participants reported that assessments are internally-mandated by team leaders, who are selected by higher management. Jasmine and Rose depend on textbooks, external books, internet, and teachers' manual as sources for assessment items and material. Rawan, Tok, and Mirjana depend on their intuition and create their questions bank, while Talen depends on previous exams samples.

#### **4.1.5.3      *Assessment Methods***

The participants reported that they use varied assessment methods. They use formal graded ones in the form of mid-term/final exams, where they assess reading, writing, listening, grammar and vocabulary. They also use less formal graded ones such as tests, quizzes, in class-assignments, project deliverables, and individual/group presentations. Mirjana said: "the less formal ones are conducted in classrooms in a relaxed atmosphere, so they do not feel anxious".

#### **4.1.5.4      *Assessment Decision-Making***

The participants reported that the team leader of every module is responsible for designing all related assessments, in addition to answer keys and rubrics. The head of the program reviews them. Mirjana commented: "once assessments are approved, they are stored and printed without having the other teachers look at them".

#### **4.1.5.5      *Assessment Design Process***

The participants mentioned that there is no specific process for exam design. They are provided with a checklist, which they have to fill regarding duration, covered chapters, number of versions, learning outcomes, and pre-set types of questions and items, which they have to follow strictly regardless of the course nature. Talen said: "we depend on our experience and tests we have seen before".

#### **4.1.5.6      *Assessment of Language Skills***

The participants explained how language skills are assessed in their context.

**Reading** is assessed in mid-term and final exams in the form of a text followed by multiple-choice and true/false questions. Rose explained; "most questions are direct ones asking for information from the text; they do not assess higher reading skills". Reading is also assessed as an in-class assignment where learners are engaged in task-based activities, which are incorporated with other skills.

**Writing** is assessed in mid-term and final exams, where students are required to write sentences, paragraphs, or essays depending on their level. Learners do other writing tasks, such as combining sentences with transitions. It is also assessed as a part of the project when learners submit their group written paragraphs for grading.

**Speaking** is assessed by asking students to prepare individual and group presentations to present in class. The presentations are based on the topics teachers provide in advance. Talen explained: "we check language, fluency, accuracy, eye contact, body language, and posture; we ask them questions during the presentation and they answer it".

**Listening** is assessed as a quiz. Assessors provide learners with a listening script followed by multiple-choice, true/false, or fill in the gaps questions.

**Grammar and Vocabulary** are assessed during in-class activities and informal mid-term/final exams as separate sections. Rawan explained: "For grammar, we use straightforward questions testing the knowledge of rules". Rose commented: "For vocabulary, learners are provided with a list of the frequent word on which they are assessed through multiple choice questions and fill in the gaps exercises".

#### **4.1.5.7      Assessment Administration**

All participants reported that the exam unit administers exams. They submit their assessments after been approved by the head of the program to the exam unit. Team leaders are involved in the printing process, after which they submit their exams envelopes to the exam unit for storage.

#### **4.1.5.8      Assessment Grading**

The participants explained that they design rubrics and answer keys internally for open-ended and closed-ended items. Rubrics are not standard; they differ from one team leader to the other. Some prefer to have holistic rubrics; others design detailed ones, which require teachers to count mistakes and penalize students for every mistake. Regarding grade allocation in answer keys, it is based on items complexity and superior's instructions.

#### **4.1.5.9      *Peer- and Self-Assessments***

The participants acknowledged that this concept is not adopted in the described context; however, some teachers provide learners with rubrics and ask them to grade oneself or others. Rose said that applying these concepts depends on having a teacher who believes in it, students' motivation, and availability of time. Rawan commented: "we do not give them the freedom to do self or peer assessment properly; it is just a set of straightforward questions to which they provide answers". Janset said that the system does not trust students to practice it because they are only concerned with grades.

#### **4.1.5.10      *Assessment Feedback***

All participants reported that learners receive feedback in the form of scores after exams. For writing and presentation; teachers do not go through rubrics in class. Jasmine explained: "we avoid giving personal feedback in class to avoid embarrassing them". Learners are invited to teachers' offices for feedback if they require so. Sherry added that feedback is never given in a supportive evaluative manner; while Janset reported: "to be honest, we do not have time".

#### **4.1.5.11      *Assessment Interpretation and Communication***

The participants explained that assessment results are posted on Banner or Moodle (soft portals for communication) in the form of scores. If learners want to see their papers, they can visit their teachers' office and ask for explanations if they need. Tok commented: "if our purpose is learning, then assessments should be communicated apart from scores and grades".

#### **4.1.5.12      *Assessment Reliability and Validity***

All participants reported that assessments designers or reviewers do not address these concepts. Mirjana explained: "we analyse item difficulty intuitively; we go through the questions; we see if there is a problem with any type; it is a personal effort". Janset commented: "they are not touching on these things; assessments are haphazardly prepared; no tutorials are given to designers on how to check validity, reliability, or scorer reliability".

#### **4.1.5.13      *Assessments Fairness and Ethical Procedures***

All participants mentioned that these concepts are not paid attention to in their context. Ethicality, in their context, is in the form of assessment storage, confidentiality, and integrity. They acknowledged that they try to check some elements informed by their experience as teachers. Mirjana said: "I try to pay attention to difficulty level, length of the exam, alignment with objectives". Jasmine said that she goes over her rubrics, share them with students, and try to make them aware of how they are assessed.

#### **4.1.6 Findings Related to Question 3.B**

*Question 3.B What are the assessors' views about their context-adopted assessment practices?*

Question 3.B seeks information about the participants' views/beliefs about their context-adopted assessment practices, which will help in understanding the fourth level of the participants' AL as per the adopted framework. Answers to the above question were reached after comparing responses provided by the eight participants on thirteen themes that resulted from the data analysis related to this question.

##### **4.1.6.1      *Beliefs about Assessment Purpose***

Jasmine, Sherry, and Talen supported the assessment purpose adopted by their context. However, the other five explained that although the purpose seems logical, the way the context manages to achieve this purpose does not match it. Tok argued that the assessments are based on testing knowledge, and practicality is almost absent. Rose explained: "students memorize things to put them in the assessments; this is demotivating and overwhelming". Janset justified: "assessments are not incorporated within learning; there is no room for fixing gaps".

##### **4.1.6.2      *Beliefs about Assessment Sources***

Seven participants supported having internally-mandated assessments, especially with diverse levels and unique specific culture. However, they reported some concerns, as they are not trained to select and design assessment material appropriately. Rose argued that there should be

detailed guidelines. Rawan believed that she should be given more freedom while sourcing assessment; she said: "When I try to go outside the box, I am told to give back". Sherry was the only one who supported using externally-mandated ones; she justified: "we do not have sufficient skills do so".

#### **4.1.6.3      *Beliefs about Assessment Methods***

All participants were satisfied with using a variety of assessments methods to address learners' preferences and reduce stress during formal ones. However, they were all against the way they were implemented. Tok explained: "from outside, it looks good that we are not adopting a monotype; but we are turning them into tests by making learners prepare them, memorize presentations, grade them using very strict detailed rubrics, and provide scores without feedback". Rose and Janset were against the number of assessments. Janset commented: "We are claiming we adopt a student-centred approach, but this is an assessment-centre approach. There are good names for assessments someone would love to hear, but they are not used as they should be; it is a testing-culture".

#### **4.1.6.4      *Beliefs about Assessment Decision-Making***

Despite being responsible for assessment practices, all participants expressed their dissatisfaction regarding excluding other teachers from assessment decision-making. Jasmine believed it should be a shared responsibility. Tok argued that assessors are overloaded and they need support from other teachers. Mirjana and Sherry added that teachers should be involved because they know what suits their learners' needs; Mirjana explained: "If teachers are involved, there will be more discussions; assessments will be more individualized and customized". All participants were against their context-adopted philosophy of assessors' selection. They believe that selection should be based on knowledge and experience with assessment; Janset commented; "the role is given haphazardly". Jasmine added: "whoever is assigned such a responsibility has to be trained". Rose added: "there is no valid unit providing guidance or auditing to ensure the validity of assessments".

#### **4.1.6.5      *Beliefs about the Assessment Design Process***

All participants were not satisfied with the assessments design process adopted in their context because there is no standard specific process to follow. Jasmine explained: "I do not know if it is the right way of doing it". Tok added: "major stuff is skipped: purpose, construct, written items, keys, and rubrics; unfortunately, we don't have time to go deep". Rose explained that even supervisors do not have the expertise to guide them with this process. Mirjana commented: "we need guidance and validation for what we do; it is a huge responsibility".

#### **4.1.6.6      *Beliefs about Assessment of Language Skills***

They expressed their dissatisfaction with the assessment of the four skills.

***For reading,*** they explained that they teach useful reading skills in class, which are not assessed, such as insertion, skimming & scanning, summarizing, analysing, ordering, and critical evaluation. They believed those reading items should be modified. Jasmine explained: "learners should be required to create something based on comprehension; these are more purposeful and reflect real authentic abilities". Tok said that he is quite happy with in-class assignments, but not with exams. He added: "We need a different kind of response because the current one might be depending on haphazard answers or choices". The other four participants commented on reading texts selection, which should take into consideration learners' levels and academic needs.

***For writing,*** six participants were against the writing assessment. They believed it is too harsh to evaluate writing based on one exam. Mirjana explained: "it is product-focused practice". Tok, Rose, and Rawan believe in the power of feedback in developing learners' writing; they believe learners should write drafts, receive feedback, and improve their writing until they reach the required level of achievement. However, they believe they might face specific challenges, such as interest in grades from students' side and feasibility of grading and giving feedback, especially with the significant number of students. Jasmine and Talen were against assessing writing in class in a processed manner. Jasmine explained: "in terms of writing process and feedback, this is a teaching process not as an assessment process".

***For speaking***, six participants were not satisfied with the speaking assessment. They do not believe asking students to present on a provided prepared topic is a useful assessment tool. Rose explained: "they memorize and come to speak in front of an audience". Janset added: "I am not satisfied with the rubric; we provide scores on nonverbal skills, subject knowledge, and preparation; how could you assess preparation". Jasmine and Talen were satisfied with presentations. Jasmine explained: "students should be able to create presentations and present to an audience". Talen added: "even if they memorize and pre-prepare the content, it will enhance their fluency".

***For listening***, seven participants expressed their dissatisfaction with the listening assessment. They were against having questions based only on multiple-choice or true-false items. They believe that this situation ignores assessing interpretation, expressing opinions, or taking notes based. On the other hand, Jasmine supported the adopted listening tests.

***For Grammar***, six participants were against assessing grammar as a separate component. Rawan explained: "not at the college level; it is not of use to them". Jasmine and Talen supported assessing grammar.

***For Vocabulary***, all participants were against assessing vocabulary as a separate component. They believe it should be incorporated with other skills. They expect to see students produce them in tasks.

#### **4.1.6.7      *Beliefs about Assessment Administration***

All participants agreed that this component is controlled and guided. Sherry explained; "when it comes to confidentiality and integrity, they are doing a good job". However, they reported dissatisfaction with their involvement in exams printing. Rawan is against treating exams as if they are sacred stuff because this is a course, where learners are supposed to learn, and teachers assist them. Janset added: "this kind of control is not for classroom assessment; it is for standardized test". Talen said: "it is better to save this effort for the assessment content".



#### **4.1.6.8      *Beliefs about Assessment Grading***

All participants reported dissatisfaction with used rubrics and keys despite attempts to develop them based on teachers' feedback and discussions. Mirjana, Jasmine, and Tok believe they need some professional guidance and standardization meetings. Janset commented: "they are either too holistic or too detailed". Those who criticize them for being holistic believe that they are unable to provide clear judgment on learners' performance. Rawan explained: "we do make them holistic because of the big number of students and management's desire to keep them simple". Those who criticize them for being too detailed are against counting mistakes. Janset argued: "We are fishing mistakes for students".

#### **4.1.6.9      *Beliefs about Peer- and Self-Assessments***

All participants believed that peer- and self-assessment should be a vital element of learners' assessments. Rose explained: "it makes students see their performance from different perspectives; learners will listen to each other more because their peers will not just penalize them for mistakes; it lowers anxiety". However, all participants had concerns about implementing them in their assessment context. Mirjana said: "When it comes to self-assessment, they are influenced by their background because their main concern is collecting scores". Tok and Sherry see the problem lies in teachers' lack of skills, tools, and techniques for applying those principles. Janset commented: "culturally, they have difficulty in accepting criticism; it may lead to some kind of harsh discussions in the class". Talen commented: "I do not trust students themselves; what kind of knowledge do they have to assess".

#### **4.1.6.10      *Beliefs about Assessment Feedback***

None of the eight participants was satisfied with the current feedback practice because it is not used efficiently. Jasmine explained: "what we provide is not personalized; it does not focus on learners' strength and weakness". Tok said: "we do not see the development". Rose explained that the challenge lies in time; there is little time allocated for giving feedback. Janset argued: "I hate seeing my students as numbers, but I do not have time". All participants believed in providing feedback for improvement that would lead to learning.

#### **4.1.6.11      *Beliefs about Assessment Interpretation and Communication***

The participants were not satisfied with the way results are communicated to learners in the form of scores. Rose believed that it is a score-oriented context. She argued: "students care only about numbers and the system is reinforcing that".

#### **4.1.6.12      *Beliefs Assessment Reliability and Validity***

None of the participants was satisfied with their context-adopted practices regarding reliability or validity because it put their assessment at risk especially that they lack the knowledge and skills that might enable them to validate their assessment or stand for it if they were criticized. Rose commented: "I do not feel comfortable. I do not trust our assessment".

#### **4.1.6.13      *Beliefs about Assessments Fairness and Ethical Procedures***

The participants were not satisfied with how these concepts are practised in their context. They explained that a lack of ethical procedures had an impact on almost everything. The assessment process is not transparent for assessors; assessment items are not addressing different styles, fixed types of questions are adopted, rubrics are not designed professionally, feedback is not appropriately provided, reliability and validity are not checked, assessors and their supervisors are not trained. Democratic practices are not applied through involving teachers and learners in the assessment. Janset commented: "No, I do not believe our assessment practices are ethical or fair". Mirjana argued: "there is no guidance to assure us that we are doing it right".

### **4.1.7 Findings Related to Question 4**

Question 4. *How can the assessors negotiate between their views about assessments and their context-adopted assessment philosophies and practices?*

The aim of question four is to elicit practical solutions to conflicts and discrepancies, which aroused between the participants' assessments beliefs and their context-adopted assessment practices. These solutions, which they proposed as a result of reflection and critical thinking, would enhance and develop their AL both on the conceptual and experiential level so that they make

use of the research experience in their current class-based assessment practices. Answers to the above question were reached after comparing responses provided by the eight participants on eleven themes representing eleven areas of conflicts, which aroused between their assessments beliefs and their context-adopted assessment practices.

#### **4.1.7.1 Assessment Purpose**

All participants reported that assessment purpose should be to infer evidence about learners' achievement in the form of real authentic use of language. Janset argued: "assessment should be conducted for learning; I get the results, and if they miss something, then I should put it back into the system and make sure that they get what I am teaching them".

#### **4.1.7.2 Assessment Decision-Making**

They all agreed that assessment decision-making should be in the hands of an assessment committee formed from experienced teachers who are well trained by experts to perform assessment tasks. Jasmine suggested selecting the members as per outstanding performance in the provided training; however, all teachers need to be aware of everything. Mirjana recommended that this committee needs to adapt what they have learned to the current context. Rose, Rawan, and Janset insisted on keeping an assessment expert for validating what the committee design at least until they are confident enough with their assessment practices. Tok suggested maintaining training periodically because new trends appear continuously.

#### **4.1.7.3 Assessment Source**

All participants mentioned that they need to depend on internally-mandated formative classroom assessments created by the assessment committee to fulfil their learning purpose. Tok explained: "we would tailor our assessments internally based on our student's needs and level". However, for evaluative purposes, four participants suggested using two external standardized assessments as entry and exit gatekeepers before the beginning of the program to assist in allocating learners in different levels and at the end of the program to exit from the program. Sherry, Rose, Rawan, and Talen refuted the ideas of external standardized assessments because they are not personalized; they

believed in internally mandated summative ones at the end of the program to validate assessor's formative assessments outcome and results.

#### **4.1.7.4 *Assessment Methods***

They all suggested removing graded midterm/final exams/quizzes and replacing them with ongoing formative classroom ones in the form of integrated task-based assignments, projects, academic activities, writing portfolios, and interviews for listening. However, they suggested that the committee should well plan those formative assessments, align them with learning objectives, and provide specific steps that enable teachers to provide unified, purposeful tasks. Jasmine explained: "we need something concrete; we need to record development in learning, document it, report it, and keep its records". They all agreed on having a summative assessment at the end of the program to validate their inferences and provide a "valid license" for exiting the program.

#### **4.1.7.5 *Assessment Design Process***

The participants mentioned that the proposed formative assessment practices would not match traditional exams design. Janset explained that teachers, along with the assessment committee guidance, need to define the required tasks, procedures, and learning outcome. For summative assessment at the end of the program, Tok and Janset suggested that the assessment committee would also do this with teachers' support by providing items to the assessment pool. If the decision is to adopt an external summative standardized assessment, then the role of the committee would be to judge the suitability of the selected assessment to learning objectives, intended learning outcomes, and the overall goal of the program; which will require amendments.

#### **4.1.7.6 *Assessment of Language Skills***

All participants agreed that learners' language skills need to be assessed through task-based assessments that require performing authentic tasks. Rose commented, "They should produce something not just answer multiple-choice and true-false questions". Janset added: "I do not believe in assessing skills separately; we can incorporate them in a project, where they use and produce all four skills without separation". Mirjana argued: "we need to move away from being test-oriented/grade-oriented to being performance-oriented". Tok added: "I

guess the problem is that we are isolating teaching from assessment; you can assess while teaching".

**Assessing Writing:** all participants agreed that it should be taught and assessed as a process. Tok said: "I would like to see writing going back and forth between teacher and students, providing feedback and students work on that". They all supported the use of portfolio for writing, where records of students' writing, provided feedback, and developed pieces are kept for learners to notice their problems and enhance them. Rawan added that students should be involved in real writing tasks. Janset added that it should be integrated with other skills, such as writing summaries or reports; teachers need to focus on different genres.

**Assessing Reading and Listening:** the participants gave different suggestions, such as giving a research assignment on specific topics. The assignment would incorporate reading, and the outcome of the task would be a spoken or a written report. They also suggested exposing learners to written/audio maps, newspaper, and messages on social media. They all believed that listening should not be assessed separately. Rose explained: "They should produce something with what they read or listen to not just answer multiple-choice and true-false questions."

**Assessing Speaking:** they suggested different speaking assessments. Rawan suggested an exit interview; Rose suggested asking learners to perform authentic spoken tasks and record their production for assessment. Mirjana and Jasmine suggested going back to participation grades that assess learners' interaction in class. Janset suggested assessing learners' speaking skill while they are giving their review on their peers work. She added: "At the end, they can present their projects to future instructors who will teach them in their academic disciplines; the instructors can assess if they are ready for their academic disciplines or not."

#### **4.1.7.7 Assessment Administration**

Three participants believed that they would not need the current administrative controlled guidelines in their new suggested assessment

practices. Rose added that teachers should not be involved in administrative practices; while the other participants believed that they need some control over summative assessments, which they will use by the end of the program. Jasmine justified: "we want to make sure that the exam is secured and done in the right way, especially with the summative ones".

#### **4.1.7.8 Assessment Grading**

They suggested that rubrics would be designed based on tasks' requirements. They will not use fixed rubrics imposed on teachers. Mirjana said: "we can have simple rubrics without overwhelming items; teachers can use error codes and feedback during grading". Janset explained: "we do not need complication; I do not want to fish for mistakes". She suggested that if learners need extra help with a particular mechanism, teachers can direct them to the tutoring and writing centre.

#### **4.1.7.9 Peer- and Self- Assessments**

All participants believe that peer- and self-assessments should be central components in their assessment process. Mirjana, Jasmine and Tok suggested providing simple checklists as an initial phase until learners get familiar with the process, then they could be provided with rubrics and error codes to assess one another work during second drafts. Janset explained: "It would be a good chance to clarify for them learning objectives, targeted skills, and required outcome".

#### **4.1.7.10 Assessment Feedback & Result Communication**

All participants believed in providing feedback and communicating their inferences to learners. Mirjana suggested submitting their tasks on Turnitin, where learners can create soft folders in which they add their soft copies and receive descriptive feedback online accompanied by error codes. She recommended using video conferences, recorded audios, and face to face interviews for providing feedback. Tok suggested referring learners who require intensive feedback to the writing centres and labs in the university. Mirjana added: "we will use the cumulative result from assessment to adjust teaching and develop learners so that they can achieve their intended objective". Jasmine suggested keeping records of all tasks, teachers' feedback, and periodic reports.

They also agreed that results from the summative tests at the end would be analysed to know learners' strength and weakness. They suggested having individual meetings with learners based on their exams results to focus on their problems.

#### **4.1.7.11 Assessment Reliability & Validity**

The participants suggested conducting meetings to discuss the validity and reliability of the formative assessments they are adopting. They also suggested collecting feedback from learners. Jasmine, Talen and Tok believed that having a summative assessment at the end would validate their formative practices, especially if it is designed with the same learning objectives in mind, and both results agree. However, they still insisted on having an expert validating their practices or a "third eye".

#### **4.1.7.12 Assessment Ethics and Fairness**

The participants believe that if they follow their suggested practices, they could judge their assessments as being fair and ethical. However, Mirjana had some doubts about the impact of teachers' subjectivity because assessment would involve only feedback with no reference to grades. Jasmine agreed: "we need a license to move them to the next levels or make them exit the program". On the other hand, the other participants did not believe that they have to justify their practices with scores. Tok said that fairness is to be fair with students and teachers through providing placement test, eliminating stressful assessments, using alternative assessments, providing learning and development opportunities through feedback and substantial assistance, and observing real language performance. He explained: "I feel that whatever I have put here now is fair and ethical to the student, it is quite satisfactory".

#### **4.1.8 Findings Related to Question 5**

*Question 5. How far have the assessors' assessment literacies been developed as a result of taking part in the research?*

The fifth research question focuses on eliciting information on how participation in the current research study contributed to developing the participants' AL both on the conceptual and experiential levels. The data derived

from this unstructured interview, though uncategorized in advance through precise questions, can provide a clear image of the impact of their participation in the research study. This knowledge will reveal one last component of assessors' assessment literacy as per the TALiP framework, which is related to their temporary dynamic assessment identity, which is formed as a result of reflection on their previous assessment education, conceptions, contextual experience, beliefs, compromises, and finally their participation in the current research stages and phases. This temporary dynamic assessment literacy is ready for more tuning, refining, and development as a result of more development, exposure, and experience with different contexts. Answers to the above question were reached after comparing responses provided by the eight participants to five themes that resulted from the data analysis.

#### **4.1.8.1 *Experience with the Open-Ended Checklist***

The eight participants agreed that going through the checklist in details had a positive impact on their AL. They all agreed that it raised their awareness of assessment concepts and practices. Mirjana explained that the checklist gave her confidence on some of her current practices, although she did them intuitively. She added that this experience changed her belief about some concepts; she illustrated: "before I thought assessment statistics is vital; now I feel we do not need that much of details". She started to believe that she needs to make radical changes, especially with classical written assessments and come up with some new techniques, which are less stressful for students. She added that the checklist made her think about details she ignored before, such as the importance of providing feedback. Tok considered the checklist a kind of self-evaluation; it was like polishing things that were not visited a long time ago. Sherry emphasised that the discussed terms and skills added to her knowledge and uncovered the lack of theoretical and experiential skills necessary for her current role. Rose found the checklist an excellent opportunity for reflection because it helped her detect her points of strength and other areas which need to be sharpened. She stressed: "I like the fact that it gave me voice about what should or should not be involved as part of an assessor's knowledge". Rawan, Janset and Talen found the checklist an excellent reference to their future assessments practices.



#### **4.1.8.2    *Experience with the Semi-Structured Interview***

They all agreed that talking about their current assessment practices, and their beliefs were a constructive reflective practice; they described this stage as an "eye-opener". Mirjana explained: "Once we started discussing everything, I started to see that certain things should change". Jasmine considered this stage the beginning of change; she emphasised: "I feel a bit bitter about the fact that we could have done things differently, but we lack knowledge". Sherry said that she felt compensated when she talked about her beliefs. Tok, Rose, and Rawan said that this practice gave them a voice. Talen added that this experience gave her some hope to start changing things; she said: "I think if I am in charge of making any assessment now, I will take into consideration all these things". Janset admitted that when she came, she had in mind not to talk about difficulties because she found them embarrassing. However, when she started talking, she felt relieved; she commented: "they were off my chest now".

#### **4.1.8.3    *Experience with the Open-Ended Report***

All participants mentioned that this was their preferred stage. Tok commented: "this utopian part made me put everything together; it also made me happy about myself that I could come up with these creative ideas". Mirjana added: "you just let me come up with my own ideal world; it made me feel I could change and I have something to share. Rose said: "It helped me put all the pieces of the puzzle together and think of practical solutions". Rawan said that this stage made her break the solid model she is trapped in and create her own satisfactory models. Janset illustrated: "I came up with an amazing project on the spot; I just in- cooperated all the energy that was charged throughout the research stages and poured it out in this stage". Talen summarised the whole situation: "It was a dream, which I lived happily". On the other hand, Sherry showed some bitterness regarding this stage; she blamed management and superiors for not providing such opportunities. She believed that this stage "touched a nerve in her" about something that she should work on. Despite the excitement the participants shared; they still had doubts about the implementation of these ideas. Mirjana commented; "I cannot really make such radical changes on my own; in our context, it is difficult to propose big changes, they can easily be rejected".

#### **4.1.8.4    *Experience with the Whole Research***

All participants talked positively about the impact of the current research on their assessment knowledge and skills. Mirjana reported that she realised that certain things are not necessary, whereas other things are fundamental. Being critical encouraged her to think of proposing some changes in the future. She also felt the need for training. She explained: "such kind of contextualized research gives a voice to teachers; this is how we learn and improve. It is not necessarily that we could find a solution". She added: "sometimes you feel relaxed if it is everybody's problem not just you, instead of blaming oneself, we start supporting each other". She emphasised that solutions to their problems will come from them. Jasmine added: "trial and error will not work anymore; participation in the research gave me such awareness". Tok commented: "This research might be an eye-opener, but if it stops at this current stage, it will be just ideas trapped in a box". Sherry added: "When I approach students with assessments next time, I will start to question my practices". Rose explained that she acquired different concepts during the research phases, which she did not know about before; she said: "I was just judging students all the time. I realised that I need to judge my practices first". Janset realised that whatever was done in terms of assessment was haphazard. She commented: "this participation shed light on my path. I am looking at assessment from a different perspective now". Rawan and Talen agreed that their current assessment practices need to be revisited because their assessment conceptions have changed.

#### **4.1.8.5    *Research Impact on Future Development in Assessment Literacy***

All eight participants would appreciate an opportunity of training in assessment on both theoretical and practical level. Mirjana explained that she needs skill-based training on grading, designing rubrics, and providing feedback. Tok said that he would like to take the research outcomes and change them into effective technical practice. Jasmine said that she would refuse any assessment role unless she is trained because it is not an easy job to do. Rose said that she would start looking for training models and suggest them to her superiors. She added that the research experience taught her to appreciate her peers. Rose emphasised: "I will start to consider ethicality of assessment in-depth; I did not use to look at its impact on students". Rawan said that she believes now that

teaching and assessment is one entity and she wants more freedom with her assessment decisions. Janset added: "I do not want all of these hours that we spent in the research to go in vain. You just opened the door, but what is next?" Talen decided to explore different sources and to depend on herself because she got interested in assessment fairness and ethicality; she said; "I need to communicate more with a community of professionals specialized in assessments".

## **4.2 Discussion of the Research Findings**

### **4.2.1 Discussion of Question 1.A Findings**

*Question 1.A How are the assessors prepared for assessment tasks?*

Analysing the findings of question 1.A showed some limitations with the participants' formal and informal preparation for their assessment tasks. I believe that the participants were not adequately prepared theoretically or practically during their pre-service or in-service phases for their current assessment tasks. This limited assessment preparation was not found only in the Kuwaiti context. Similar results were found in studies conducted in Canada, China, and Korea (Deluca & McEwen, 2007; Jin, 2010; Lam, 2015). They found that assessment training was missed by teachers either because assessment courses were offered as electives or because their training was free from any assessment component and they were mainly focused on methodology (Greenberg & Walsh, 2012). Lack of assessment preparation threatens the participants' eligibility for their assessment roles (Sato, Wei & Darling-Hammond, 2008). Assessment support provided for teachers can enhance their APs as reported in other empirical research studies (Mahapatra, 2016; Montee, Bach, Donovan, & Thompson, 2013; Nier, Donovan, & Malone, 2013; Walters, 2010). Lack of training in assessment affects TAL, as studies reported a positive relationship between assessment training and TAL (Levy-Vered & Alhija, 2015; Quilter & Gallini, 2000).

It was also inferred that the participants' performance on their assessment tasks depended on self-professional development in the form of free-reading, online tutorials, looking up rubrics online, joining professional organizations, or

discussion with colleagues. This finding resonates with findings of AL studies that specified some channels, to which teachers yield as alternatives to formal channels, such as online learning resources (Fan, Wang & Wang, 2011), support from within the workplace (Lukin, Bandalos, Eckhout, & Mickelson, 2004), and daily classroom practices (Smith, 2011). However, all these studies emphasised that seeking these resources does not deny the need for sustainable assessment training for assessors to be able to perform their assessment tasks competently.

#### **4.2.2 Discussion of Question 1.B Findings**

*Question 1.B How are the assessors involved in assessment tasks in their teaching context?*

Analysing the findings of question 1.B, I believe the participants are intensively involved in assessment tasks. Their involvement in assessment decisions in the described context is of high stake. They are almost responsible for all assessment decision-making and practices, including methods, design, administration, grading, communication, and announcement. These intensive roles require them to be highly assessment literate teachers (Taylor, 2013). However, their selection for the mentioned assessment role was not based on their knowledge of and skill in assessment; it was done based on administrative criteria that did not take into consideration their assessment competence or literacy. The selection criteria adopted in the described context does not match the criteria or standards set for teachers to perform assessment tasks. Teachers should be chosen based on their high competence at and confidence with assessment knowledge and skills (Brookhart, 2011). Coombe, Troudi, & Al-Hamly (2012) emphasised that without a higher level of TAL, teachers will not be able to help students attain higher levels of academic achievements. When teachers lack assessment preparation, they make erroneous decisions, and students will suffer from negative consequences (Purpura, 2016).

### **4.2.3 Discussion of Question 2.A Findings**

*Question 2.A How confident are the assessors with the assessment knowledge base suggested by the TALiP framework?*

Analysing the findings of question 2.A, it is inferred that at least seven of the eight participants lack confidence with the assessment knowledge base proposed by the TALiP framework. First, this lack of knowledge is attributed to limited theoretical and technical preparation, which the participant's themselves reported as an answer to question 1. A. This finding is supported in AL literature, as it is reported that there is a positive correlation between teacher training in assessment and their acquired knowledge and skills (Muhapatra, 2016; Montee, Bach, Donovan, & Thompson, 2013; Nier, Donovan, & Malone, 2013; Walters, 2010). Second, the consensus that the participants showed while reporting their confidence level with the components and the sub-components of the assessment knowledge base is attributed to the nature of their group. They were a relatively homogenous group, which shared common features. They received similar contextual methodological training. They were guided by the same contextual assessment philosophies. They spent more than five years as a close-knit group performing the same required tasks and abiding by the same assessment contextual philosophies with very limited room for assessment professional development.

Seven participants showed limited confidence with almost most of the components and sub-components of the proposed assessment knowledge base except for some familiarity with components that are related to teaching methodology, learning theories, general feedback principles, and some general ethical considerations. However, they showed limited confidence with all components that require assessment knowledge and skills as discussed thoroughly in the below section.

The participants' confidence with the sub-components of disciplinary and pedagogical knowledge is directly related to their methodological training, with which they were familiar. This finding supports the direct relation between teachers' training and competence in trained topics (Yastibas & Takkac, 2018). It also emphasises other research findings, which showed that teaching

methodology is one of the dominant components in teacher education and that AL is the least represented component in teacher education (Abell & Siegel, 2011).

The participants' limited confidence with the theoretical knowledge of assessment purpose, content and methods is attributed to lack of assessment training. Limited confidence with these components makes assessors unable to defend their assessment practices against imposed policies (Shohamy et al., 2017); it also controls learners' opportunities for further education (Fulcher, 2010). The participants' limited confidence with these theoretical components contradicts emphasis given to them in AL literature, as they affect the quality and validity of designed or selected assessments (Brookhart, 2011; Weir, 2005). Moreover, the participants' limited confidence with skills associated with this knowledge is also attributed to lack of training and guidance. Lack of these skills can lead to unjustifiable decisions about assessment and put assessment validity under threat (Popham, 2011).

The participants' limited confidence with the theoretical knowledge of assessment grading is also attributed to lack of training. This finding contradicts attention given to grading knowledge in AL standards literature (AFT, NCME, & NEA, 1990; Brookhart, 2011; JCSEE, 2015). Lack of grading knowledge leads to invalid assessment and grading practices (Brookhart, 1999). Brookhart (2011) believed that limited confidence with grading knowledge could affect teachers' ability to infer meaning from learners' results. In addition, the participants' medium confidence with some grading skills associated with rubric design is also critical, as using intuition and experience with teaching as the only base for designing rubrics does not deny the need for sustainable assessment training (Smith, 2011). Teachers should be able to justify their grading practices informed by solid knowledge (Fulcher & Davidson, 2007).

The participants' confidence with some theoretical knowledge related to assessment feedback is conceptual one guided by their teaching skill not by solid academic training, which is highly recommended in AL literature (Brookhart, 2011; Cumming, 2009). Limited confidence with knowledge related to kinds, types, and mechanism of performing feedback tasks is not supported in AL

literature, as this knowledge is essential and fundamental in performing feedback tasks (Cumming, 2009). Writing clear and useful feedback requires training (Lee, 2009; Fulcher, 2010). The participants' confidence with some practical skills related to assessment feedback, such as setting goals, identifying learning gaps, and providing criteria of success for learners is questioned because they reported it is guided by their teaching experience. Their limited confidence with intervening techniques, scaffolding exercises, recording techniques, or using internet resources is not supported in AL literature (Brookhart, 2011; Cumming, 2009; Fulcher 2010; Hattie & Timperley, 2007; Lantolf, 2009). It can be inferred that the participant's knowledge with practical feedback skills is superficial. Performing feedback tasks requires more professional training (Lee, 2009)

The participants' limited confidence with the theoretical components of peer- and self-assessments contradicts Brookhart's (2011) emphasis on TAL in self-and peers-assessment principles and strategies because they enhance their roles in assisting learners in interpreting assessment results, tracking their learning, communicating about their learning, and planning next steps in their learning. Understanding self-and peer assessment concepts helps assessors select and use formative assessment strategies and work with students appropriately (Moss & Brookhart, 2009). Moreover, the participants' limited confidence with some practical aspects of peer and self-assessments does not align with AL literature, which advocates enhancing teachers' skills in this topic (Furtak et al., 2008; Torrance & Pryor, 1998). Teachers are expected to support learners' feelings of self-efficacy and control (Bandura, 1997; Ryan & Deci, 2000) and self-regulation (Butler & Winne, 1995) through involving them in assessing themselves and others.

The participants' limited confidence with the theoretical knowledge related to assessment interpretation and communication does not resonate with Brookhart's (2011) recommendation, which emphasises teachers' ability in interpreting evidence derived from norm- and criterion-referenced assessments, in addition to ensuring that their interpretations have sufficient evidence, which would assist them in improving learning. Their limited confidence with skills related to this knowledge, especially those related to statistical procedures, such as item difficulty, discrimination, reliability, and validity is not supported in AL

literature. Stiggins (2008) recommended that teachers should be able to infer learners' performance, areas of weakness, and accordingly justify obtained results while communicating them to stakeholders.

The participants' limited confidence with the theoretical knowledge of assessment ethics due to lack of training does not resonate with Brookhart's (2011) recommendation, which emphasises that teachers should understand legal and ethical responsibilities associated with their assessments and their impact on learners. In addition, their reported limited confidence with skills associated with this knowledge is also not supported in AL literature. They are required to work towards equity, non-discrimination, inclusion, and social justice (Tierney, 2013). Lack of training in ethics skills can affect their assessment negatively (Lynch & Shaw, 2005).

Exploring the participants' confidence with the assessment knowledge base showed limited confidence with most of its components. An awareness of formal systematic codified knowledge of assessment is a necessary and essential component of LTAL (Fulcher, 2012). As per the adopted TALiP, framework, assessment knowledge base with its theoretical and practical components is a fundamental primary component of LTAL without which teachers cannot engage with assessment at a deeper level. How far other levels of LTAL will be affected by limitations in this primary level will be discussed in the below sections.

#### **4.2.4 Discussion of Question 2.B Findings**

*Question 2.B What are the assessors' views regarding their need for different components of assessment knowledge base suggested by the TALiP framework?*

Analysing the findings of question 2.B, it is inferred that the participants showed the need for almost all theoretical and practical aspects of all components of the assessment knowledge base. The following section presents the relation between their shared views and the literature findings.



First, their need for the components of disciplinary and pedagogical knowledge is emphasised in AL literature (Brindley, 2001; Davies, 2008; Inbar-Lourie, 2008a), as it is urged that this knowledge guide assessments decisions related to the selection of tools, items, and content (Abell & Siegel 2011). Meanwhile, it is emphasised that assessors' knowledge of teaching theories has a positive impact on their assessment practices (Ball, Thames, & Phelps, 2008).

Second, the participants' need for nearly all theoretical and practical components of knowledge of assessment purpose, content and methods is supported in AL literature (Brookhart, 2011; Combee et al. 2012; Fulcher, 2010; JCSEE, 2015; Shohamy et al., 2017). Their concern about the negative washback of their assessment is supported in Hughes (2003) and Weir (2005). The justification they provided for their need for these components to produce fair, ethical assessment practices was also emphasised in Davidson & Lynch (2002). The participants' needs resonate with TAL standards proposed in Davies (2008), Fulcher (2012), Inbar-Lourie (2008a), and Popham (2011), who emphasise acquiring these skills as significant components of TAL. Third, their need for knowledge of assessment grading is supported in Brookhart (2011), as she believed that knowledge about constructing scoring schemes that quantify learners' performance is unavoidable for making useful inferences, which would lead to improved learning. Their need for grading sub-skills is supported in AL literature (Campbell & Collins, 2007; Fulcher & Davidson, 2007).

Similarly, their need for knowledge of assessment feedback is also supported in AL literature (AFT, NCME, & NEA, 1990; Brookhart, 2011; JCSEE, 2015). Their justification resonates with research findings, which emphasised learners' right to know the purpose behind their learning (Hattie & Timperley, 2007) and their strength and weakness (Shute, 2008). Their need for practical training on how to perform these skill is also supported in AL literature (Cumming, 2009; Leung, 2004; Leung & Scott, 2009; Shepard, 2007). Moreover, the participants' need for nearly all components of knowledge of assessment interpretation and communication is emphasised in Hughes (2003) and Popham (2011), who necessitate interpretations of both criterion- and norm-referenced results for language teachers to ensure assessment validity and fairness. Finally, the participants' consensus regarding their need for nearly all components of

knowledge assessment ethic echoes Stoyhoff's (2008) view, who believed that this knowledge minimises the negative consequences of assessments.

Although the participants supported their needs for different components of the assessment knowledge base by valid reasons that resonated with findings of different research studies conducted on TAL needs, I argue that they did not manage to filter their needs as per their conceptions and assessment needs. This filtering stage was one aim of exploring their need for the proposed assessment knowledge base. I believe the number of needs and justifications the participants expressed was exaggerated. In many cases, teachers express a need for "everything" rather than being specific about their needs (Deluca & Klinger, 2010). This finding shows that teachers' perception of their assessment training needs may not be what they want (Xu & brown, 2016). I believe that if the participants were more confident with the assessment knowledge base and more aware of its components and its skills, they would have been more specific and precise about their needs and its relevance to their context (Howley, Henning, Gilla, & Weade, 2013). This argument is supported by the participants' attitude, which was observed while reporting their need for the various components. They continuously associated their needs with their limited assessment training and insisted on receiving technical professional training through accredited useful resources as a source for fulfilling their needs.

#### **4.2.5 Discussion of Question 3.A Findings**

*Question 3. A How are assessments practised in the assessors' teaching context?*

Analysing the findings of question 3.A, I argue that the assessment philosophy underpinning the described context APs is influenced by an assessment philosophy focusing on assessment of learning to evaluate learners' achievement of pre-set learning objectives through different summative assessments. Although the context adopts different assessment methods, they are used in a summative manner for scores collection and not for formative reasons. All assessments are internally-mandated by selected teachers, who are assigned assessments and rubrics design tasks. Assessment design does not follow defined procedures or steps. Feedback on assessment design is provided

only on the layout with no reference to objectives, constructs, test specifications, item design, or content of the assessments. Respective language skills are assessed through close-ended questions, while the productive ones are assessed through prepared topics that are memorized and performed during exams. Assessments are graded through controlled rubrics prepared by the assessors. Assessment results are announced in the form of scores; no evaluative or descriptive feedback is provided to learners for development reasons. Peer- and self-assessment concepts are not adopted in the described context. Assessment validity, reliability, or fairness are not negotiated or checked in the described context.

As per explored literature, this adopted philosophy is not supported in EPP programs where the emphasis should be on promoting learners' performance (Richards & Renandya, 2002) not on gathering information about learners' achievement for judgment (Shohamy, 2001; Lee, 2011). The adopted assessments methods do not include authentic performance tasks that require learners to use language to perform academic tasks (Cameron, Tate, Macnaughton, & Politano, 1998). Depending on only provided scores and lack of feedback eliminate any chance for learners' development and learning from their assessment experience (Sadler, 1989). The participants' description of their context-adopted practices echoes and validates my observation and the contextual rationale behind conducting the current research study.

Therefore, I believe that question 3.A succeeded in revealing another component of LTAL as per the TALIP framework, which is their context-adopted assessment philosophies and practices. These philosophies set boundaries for their assessment practices in terms of what they should do (Gu, 2014). I believe that the participants' limited confidence with the assessment knowledge base and limited assessment education reported in previous discussions reinforced their confrontation to their context boundaries (Wyatt-Smith et al., 2010; Xu & Liu, 2009).

#### **4.2.6 Discussion of Question 3.B Findings**

*Question 3.B What are the assessors' views about their context-adopted assessment practices?*

Analysing question 3.B findings revealed the participants' conceptual awareness of adequate APs, complexities, and threats. This conceptual awareness was strong enough to enable them to reflect on their contextual practices competently. They showed a solid awareness of contemporary reforming assessment practices in assessment literature. I believe this awareness is informed by their experience, daily experience with teaching and assessment, and their reflection on surrounding practices. I argue that if the participants were more confident with the assessment knowledge base, and they were academically and practically prepared, they would have revealed more expanded knowledge and more critical views.

Exploring teachers' beliefs about their context-adopted APs showed discrepancies between their conceptions and the boundaries imposed upon them within their context. The tighter the boundaries, the less space there is for professional autonomy. Tensions arise for teachers when they have less autonomy (Fleer, 2015; Forsberg & Wermke, 2012), which is the case in the described context since this tension resulted in dissatisfaction with almost all APs.

The participants' dissatisfaction with most of their context-adopted assessment practices is supported in assessment literature. For example, their dissatisfaction with their context-adopted assessment purpose, which focuses on collecting information about learners' achievement through scores for evaluation not for providing learners with learning experience was reported in different contexts by other EFL teacher in similar studies (Efeotor, 2017; Richards & Renandya, 2002). Their dissatisfaction with their selection for their assessment roles, their concern about their assessment knowledge base, and their fear of the negative impact of their limited training and inadequate competence in sourcing, selecting, and designing assessment material resonate with Purpura's (2016) concerns. Purpura believes that when assessors make erroneous decisions due to lack of knowledge, learners suffer from negative consequences. The participants' dissatisfaction with excluding other teachers from assessment decision making because it deprived them of their right to share their ideas was reported by EFL teachers working in similar EPP in United Arab of Emirates and Kuwait (Troudi et al. (2009).

Moreover, the participants' dissatisfaction with the way multi assessment tools were miss-used in their context was emphasised in (Bachman, 2005). Bachman warned against ignoring the informative role of these tools and using them for scores collection, not for formative reasons. Coombe et al. (2012) believed that if multi-measure assessments did not involve authentic assessment tasks, provide positive feedback, or improve instructions, they lose their value and fail in achieving the aim behind their use. Coombe's concerns resonate with the participants' views. They believed that the miss-use of the various assessment tools negatively affected the learners and made them work for test scores instead of using these multi-tools as learning opportunities and means through which they express themselves differently away from stressful conditions accompanying official tests.

Similarly, their dissatisfaction with the unstandardized and unguided assessment design process adopted in their context was emphasised in assessment literature. The context-adopted design process contradicts with assessment process design recommended in Davidson & Lynch (2002), Fulcher & Davidson (2007), and O' Sullivan (2011). Meanwhile, the participants' disbelief in the product-based assessment approach adopted by their context in assessing language skills and their belief in task-based approach to language skills assessments resonate with Shohamy's et al. (2017) ideas regarding involving authentic task-based items in assessments. They believe that task-based assessments reveal a real image of learners' competency instead of assessing pre-determined knowledge as adopted by product-based assessments. The participants' dissatisfaction with reading and listening assessments, because they encouraged haphazard answers by focusing on specific closed-ended response items, resonates with Weir's (2005) warning against such items.

Finally, the participants' dissatisfaction with the grading procedures and result communication procedures adopted in their context, which focused on producing scores rather than feedback for learners, aligns with feedback principles advocated by Cumming (2009) & Leung & Scott (2009). The participants' support to peer-and self-assessment, which were ignored in their context, is advocated by Shepard (2007); however, they associated them with appropriate training for teachers to be able to implement them properly to gain

their benefits. Neglecting peer-and self-assessments deprive learners of the chance to notice the gap in their learning process (Black et al., 2002, 2004). In addition, the participants' dissatisfaction with lack of attention paid to assessment validity, reliability, and fairness in their context echoes Popham's (2019) argument regarding the importance of applying these notions in language assessments.

#### **4.2.7 Discussion of Question 4 Findings**

*Question 4. How can the assessors negotiate between their views about assessments and their context-adopted assessment philosophies and practices?*

The participants succeeded in providing compromises, which resonate with recommendations of research conducted on AL (McMillan, 2003). They suggested adopting an assessment philosophy, which focuses on integrating assessment with learning by focusing on what learners can produce with their acquired knowledge, not their ability to reproduce what they learned. They also suggested providing constructive descriptive feedback on learners' performance in the form of written reports, taped videos/audios, soft portals, or meetings to enhance their learning and adjust both learning and teaching. This approach resonates with Brookhart's (2009) emphasis on this type of assessment philosophy as a learning agent rather than being an evaluative one. They suggested depending on internally-mandated formative assessments using different assessment tools such as, task-based assignments, projects, academic activities, portfolios for writing, interviews for listening, which can assist them in achieving their assessment philosophy. This view aligns with suggestions by Airasian (2005), Bachman (2005), Coombe et al. (2012), and McMillan (2007), as these tools enable learners to produce authentic performance, which can help assessors infer fair decisions. They supported depending on task-based activities to teach and assess learners' integrated skills. They also suggested incorporating peer and self-assessments as central tools for assisting learners in noticing their weakness and improve independently (Moss & Brookhart, 2009).

Although the participants provided solutions that can solve problems, which they reported in their context, they emphasised that they need guidance and assistance to validate their assessment practices through a professional

committee that can provide them with ongoing training and feedback on their assessment practices. This need is attributed to their limited preparation and confidence with the assessment knowledge base, which affected all other components of their AL. This need validates Xu & Brown's (2016) claim regarding the value of acquiring assessment knowledge base because its absence can deprive teachers of engaging with assessment at a deeper level. The participants' suggestion echoes Yastibas & Takkac's (2018) call for ongoing support required for teachers in their assessment tasks. It is reported that adopting training practices would make the participants satisfied with their job because they will be able to provide fair opportunities for learners, which makes their practices ethical. This view is also supported in Tierney (2013) & Stoyhoff (2008).

Despite the participants' constructive solutions, it is found that they still hold on to summative assessments as tools for validation of their formative assessment practices. The participants' fear of social pressure and blame, although they are convinced with their suggestions, makes them suggest exist summative tests along with all the formative assessment practices they suggested. This finding supports Shohamy's et al. (2017) claim about the power tests have on teachers and students, which stands as an obstacle slowing down implementing innovative, beneficial practices (Inbar-Lourie, 2008a; Shepard, 2000).

#### **4.2.8 Discussion of Question 5 Findings**

*Question 5. How far have EFL teachers'/assessors' assessment literacies been developed as a result of taking part in the research?*

Analysing question five findings, it is inferred that the participants' AL has developed to a certain extent as a result of taking part in all stages of the current study at least on the conceptual level.

As reported, taking part in the research raised the participants' awareness of assessment concepts and practices and made them confident with some APs they used to practice intuitively. It also changed some rooted assessment beliefs and provided them with new ideas that could be implemented in their contexts. In addition, it provided them with an opportunity to self-evaluate and revise their

practices. Most importantly, it uncovered a lack of specific knowledge and skills and gave them a voice to utter what they need regarding all proposed concepts and skills. These research gains uttered by the participants echo Xu & Brown's (2016) expected benefits, which teachers would feel as a result of going through stages of the proposed model, as it is not meant only for exploring their AL as much as it is meant for developing their AL.

Providing the participants with the opportunity to talk about their context assessment practices and their beliefs about them uncovered hidden realities, made them realise uncertainties and problems, and encouraged them to think of steps to change and develop practices. They started to feel the discrepancies between their practices and beliefs, and they put their hand on reasons for their dissatisfaction with their assessment practices. This finding resonates with ideas proposed by Fleer, (2015) and Forsberg & Wermke (2012), who believed that such research practices are essential because it can show the kind of incongruence that might arise between teachers' conceptions and the boundaries imposed upon them within their context.

Providing the participants with an opportunity to suggest ideas and solutions had a positive impact on them, as it gave them confidence in their conceptions and experience when they saw the outcome of their ideas. It gave them hope for better practices, and it encouraged them to think of sharing their ideas with management and take further steps. It made them realise how things could be done in a better way. These findings echo Xu & Brown's (2016) ideas about the benefit of problematising practices and reflecting on them through solutions.

Moreover, taking part in the study made the participants realise the ethical need for more professional training to guarantee fair practices proposed to learners. They realised the importance of taking their gains from the research to their daily assessment decisions, which they would revisit. These benefits are supported in AL literature. Xu & Brown (2016) believe that engaging teachers in professional conversations about their APs may lead them to make subsequent changes in their day-to-day assessment practices.



Finally, I believe that involving the participants in the research had an impact on their professional development plans. They all decided to consider training opportunities in assessments through their teaching context, assessment communities, or self-studies. They started to be interested in areas such as assessment ethics, philosophies, grading, and feedback. They showed interest in the current research outcome, as they believed that solution to their problems would come from such contextualized studies. The participants' report is emphasised in Wyatt-Smith et al. (2010), who believed that such research practices could raise teachers' awareness, empower them with autonomy, resources, and voice, in addition to assisting them in reclaiming their ownership of assessment practices.

Answer to question five showed different gains and developments the participants themselves reported as a result of taking part in the study. I believe that the study succeeded in uncovering their AL and managed to raise their awareness regarding their AL, context-adopted adopted APs, assessment complexities, and potential reform that could happen in their practices and their context-adopted philosophies. However, these gains should not deny the urgent need for professional technical and theoretical training in assessment on the job. This training should take into consideration their contextual boundaries, which they reported in the current study.

## **Conclusion**

Analysing the findings of the five research questions, I believe the proposed answers succeeded in fulfilling the research purpose, which was focused around exploring and possibly developing TALiP as per the adopted TALiP framework. I believe going through these stages contributed to identifying the participants' assessment identity, which was one main concern of the current study, in addition to uncovering the context-adopted assessment philosophies. I believe this identification shed light on the source of the observed assessment problem in the described context, which was the primary rationale behind conducting the current study.

## **Chapter Five: Conclusion**

### **5.1 Summary of the Findings**

The current study was conducted with the aim of understanding, conceptualizing, and possibly developing eight EFL teachers'/assessors' AL in an EPP in Kuwait. The arguably limited APs observed in the described context was the primary rationale behind conducting the study. Learners and other EPP teachers, who were not involved in APs were not satisfied with the adopted assessments. The study was after exploring the assessors' AL and the context-adopted assessment philosophies to identify the source of the observed assessment problem and if it was the result of limited TAL, context-adopted assessment philosophies, or both. Being aware of the source of unsatisfactory APs can guide improvement and development in the described context.

Moreover, limited APs was observed in similar programmes in Kuwait. It was also noticed that APs related to EAP was unrepresented in academic references, journals, and research studies. Therefore I argued that conducting the current study would give a clear understanding of the observed problem, contribute in filling a gap in the EAP assessment literature, and finally attribute to the overall research conducted on TAL. The study was underpinned by a socio-cultural view while exploring TAL, one that took into consideration all variables that affect TAL not just assessment knowledge base in the form of theoretical knowledge and a set of skills that teachers/assessors are required to acquire. Therefore the study adopted the TALiP framework proposed by Xu & Brown (2016), as a theoretical framework to explore the assessors' AL.

TALiP framework is based on the assumption that TAL consists of different interrelated levels that form their overall assessment identity as assessors. The fundamental component of this literacy is assessment knowledge base, which consists of theoretical and practical components that assessors need to acquire before engaging in APs; it will guide their assessment conceptions and practices. The assessors' beliefs filter this knowledge to decide what is needed or not needed from the assessment knowledge base as per their assessment requirements, preferences, and views. TAL is also affected by the contextual assessment boundaries. These boundaries impose on teachers certain

philosophies and practices, to which they need to respond through compromises between them and their own assessment beliefs. Assessors reach these compromises through reflection, critical suggestions, and engagement in professional dialogues with colleagues. These reflective practices enable them to reach an assessment identity that can enable them to meet their required assessment tasks competently and ethically to propose fair assessment practices for their learners.

Informed by the TALiP model, the research sought answers about the assessors' assessment education, assessment roles, and selection criteria. It also questioned the participants' confidence level with the assessment knowledge base proposed by the TALiP framework and their needs for its components. Besides, it questioned the participants' context-adopted assessment philosophies, and their beliefs about these philosophies, and how they compromise between their beliefs and those imposed philosophies. Finally, the research explored the conceptual development that happened to the assessors as a result of taking part in the current study.

I believe answers of the research questions succeeded in addressing all the components of TAL, and therefore gave a clear understanding of the assessors' AL, which was the primary purpose of the research. Exploring the participants' AL, it was found that some of its components reflected competency, whereas others did not. It was found that the participants were not sufficiently prepared theoretically or practically during their pre-service or in-service preparation phases for their intensive assessment tasks. This gap impacted negatively on the first component of their assessment literacy, which is their assessment knowledge base. The participants were not confident with most of the components of the knowledge base proposed by the TALiP framework, especially those related to assessment design, grading, feedback, peer- and self-assessment, results' interpretation and communication. They showed some confidence with some concepts, which were related to teaching such as, knowledge of learning and teaching theories, disciplinary knowledge, general principles of feedback, and other general principles related to assessment ethics, which could be acquired on the job. However, they showed limited confidence with almost all practical skills and techniques, which required academic input and

training such as, writing specification, writing reliable items, designing grading rubrics, using feedback techniques and exercises, statistical calculation related to validity and reliability, statistical calculations related to results interpretations, and statistical principals related to standardized tests.

Moreover, exploring the assessors' conceptions about their need for assessment knowledge, it was found that the participants showed the need for almost all theoretical and practical aspects of the proposed assessment knowledge base. They did not manage to filter their needs as per their conceptions and assessment needs, which was one aim of exploring their needs for the proposed assessment knowledge base. I believe that if the participants were more confident with the assessment knowledge base and more aware of its components and skills, they would have been more specific and precise about their needs and its relevance to their context. They continuously associated their shown intensive needs for almost all components of assessment knowledge base with their limited assessment training, and they insisted on receiving technical professional training through accredited useful resources as a source for fulfilling their needs.

In contrast, exploring their contextual macro- and micro-features and their beliefs about these contextual boundaries, which constitute another level of their AL, it was found that they showed awareness of assessment complexities, boundaries, and uncertainties in their teaching context. They were able to distinguish appropriate and inappropriate APs and their washback on learning. Their shared beliefs about assessment resonate with recommended APs in AL literature. Similarly, their shared solutions to problems related to their context-adopted assessment practices reflected their ability to reflect, critique, and find solutions. However, this awareness was informed by their experience, daily experience with teaching and assessment, and their reflection on surrounding practices. I argue that if the participants were more confident with the assessment knowledge base, and they were academically and practically prepared, they would have revealed more expanded knowledge and more critical views. Moreover, exploring this level of their AL showed discrepancies between their conceptions about APs and their context-adopted assessment philosophies, which were imposed upon them. This finding shed light on their limited contextual

assessment philosophies, which might have contributed to the observed assessment problem along with their limited confidence with the assessment knowledge base that should have guided their APs.

Finally, the participants' ability to engage in a professional dialogue about their strength, weakness, and future required development in assessment revealed another level of their literacy, which the study sought to reach. Although the participants showed some conceptual gains and development concerning AL, which they reported as a result of taking part in the study, I believe that LTAL development in the current situation requires different development protocol, which targets developing their assessment knowledge, skills, and means of transferring this developed knowledge to their current APs through professional expertise and trainers.

I believe that the study succeeded in uncovering the assessors' AL and managed to raise their awareness regarding their AL, context-adopted adopted APs, assessment complexities, and potential reforms that could happen in their practices and their context-adopted philosophies. However, these gains should not deny the urgent need for training the participants and developing their theoretical and practical assessment knowledge base. I believe that limitations within this fundamental component affected all the other components of the participants' AL. Exploring the assessors' AL, I can strongly argue that the observed assessment problem is attributed to both the participants' limited assessment knowledge and the assessment philosophies adopted by the described context, which affected the adopted APs negatively.

## **5.2 Research Implications**

Arguably, the findings above suggest actions on different levels. The observed assessment problem in the described context needs to be addressed both on the assessors' level and the contextual level. As a member of the academic operations unit in the described context, I intend to share the findings of the current study with the academic committee. The detected limitations in the assessors' assessment knowledge need to be treated through professional training on campus as required by the participants. Areas of weakness and the

participants' needs have to be addressed, as they cannot go on performing their assessment tasks with the same approach after the awareness they experienced as a result of taking part in the study. I intend to propose a small-scale project that aims at addressing detected AL gaps by providing professional development opportunities to faculty. These opportunities could take the form of academic lectures and technical workshops provided by professional trainers in the assessment field, training opportunities in assessment training centres, attendance of assessment conferences, or attending Webinar sessions. This training should be presented to all faculty involved in assessment decision making taking into consideration that faculty members are required annually to suggest and request professional development opportunities funded by the described context. Therefore, I argue that we should seize this funded professional development opportunity in addressing problems that were revealed as a result of taking part in the current exploratory practice.

Being responsible for the enhancement of academic policies in the described context, I intend to suggest improvements to the adopted assessment philosophy in the EPP concerning the assessment purpose, design process, methods, grading, feedback, and result communication. Raising the assessors' concerns and doubts during the quarterly academic meeting and suggesting reform in these areas guided by the explored literature, assessors' recommendation, and similar studies recommendations could be one step further in developing the programme's assessment philosophies and practices. Therefore, I plan to present a report to the committee accompanied by a suggested project plan to implement the recommended changes.

### **5.3 Recommendations**

Based on the finding of the research revealed ahead, I recommend conducting a comparative research study involving other EPP, which are few in Kuwait, in similar exploratory practice using the same approach. Comparing TAL in these programs could help researchers reach general boundaries or parameters of EAP assessors' AL required for this kind of programmes. These parameters would guide training courses targeting teachers/assessors working in these programs. EAP practitioners need a practical guide on how to perform

their assessment tasks, so working on collaborative projects between EPPs and educational practitioners on continuous training programs would help bridge the gap between theories and practice and would involve assessors in their learning process. The outcome of such projects would fill the gap in EAP assessment literature and provide solid practical knowledge and skills for EPP assessors because they are rooted in and derived from real educational contexts with all its complexities and variables.

Moreover, I recommend including theoretical and practical assessment training as a central component of teacher education and the main component of their teaching license. Assessment courses for teachers should not be kept as electives or be excluded from education or training programmes. Moreover, I recommend that assessment roles should be given to language teachers based on their proven useful knowledge and skills in assessment, whether through accredited bodies, certifications, or skill assessment. Designing, grading, and interpreting learners' assessments should be in safe hands, as it is not only a professional duty; it is an ethical one that is meant to implement justice and equality among learners.

## **5.4 Research Contribution to Knowledge**

I believe that adopting the TALiP model as a theoretical framework for researching the assessors' AL in the current study was a key step in bridging the gap between theory and practice and a real contribution to AL knowledge. Operationalizing the TALiP model in the study was an opportunity to evaluate its feasibility and practicality in exploring TAL in a real education context. Theoretically, I found it an ideal framework for exploring TAL because it included almost all TAL components suggested in AL literature; however, changing the model into an evaluative practical tool to explore or develop TAL was a real challenge. I found it challenging to design different research tools to explore the six components of the TALiP model to be able to form a holistic integrated understanding of TAL. The assessment knowledge base proposed by the model was detailed and inclusive. Although it was challenging to design a research tool that can explore all components of the knowledge base, but the participants found it beneficial and educative. Meanwhile, exploring the participants' beliefs about

assessment in general and their context-adopted assessment practices revealed rich AL knowledge that was compatible with theoretical knowledge in assessment references because it was derived from real contextual situations. Moreover, including the assessors' reflective practices, personal professional development, and colleagues' professional dialogue as major components of the assessors' AL was a valuable contribution of the TALiP model because it added a critical dimension to TAL and made the participants reveal unexpected genuine contextualized assessment ideas and solutions to challenges they face in their current context. However, the only component of the TALiP model that was not applicable to the research participants was exploring their needs for the different components of the assessment knowledge base. In the current study, the participants were not confident with the proposed assessment knowledge base; they were not able to select what exactly they needed, and therefore expressed their need for all components. I argue that this component of TAL could be explored only if the participants received valid assessment education. In this case, the assessors would be able to judge what they need guided by their education, beliefs, and contextual experience. This contributed knowledge about their needs would reveal key information about their AL, which is their ability to select what they need from the intensive assessment knowledge base available for them. Therefore, I recommend using the TALiP model in other research studies in different contexts as a well-established framework upon which researchers can build their research methodology.

Moreover, I arguably consider the checklist (appendix 10.2), which I designed to explore the participants' knowledge base, a valid tool for exploring and developing TAL in similar contexts. It is a long detailed document that includes TAL components suggested in AL literature. It involves theoretical and practical assessment knowledge with related sub-skills and components. The checklist can influence assessors' professional practises. EAP assessors can use the checklist as a learner-centred educational tool; it can direct them towards different components they need to take into consideration while performing their assessment tasks, or it can be used as a self-evaluation document, which can direct them to topics they need to explore if they were not familiar with its components. The checklist could be used in teachers' education courses. It could be used for designing an assessment guide for EAP teachers if it is accompanied



by details, theoretical explanation, and practical examples that can clarify how the sub-skills could be practically implemented. Such education courses could take the form of short seminars, workshops, or online courses, such as Massive Open Online Courses (MOOCs). Researchers conducting research on TAL can adopt and adapt the checklist and use it for exploratory or development reasons.

Finally, I believe that the research findings would have a positive impact on EAP assessors' understanding of their literacy and their future assessment development decisions. Awareness, concerns, and dissatisfactions shared through the current research would act as an eye-opener for EAP practitioners to take further reformation steps based on the research findings. Moreover, the findings of the study may guide assessment training bodies on assessment training courses for EAP practitioners and inform them regarding the needed content of these courses. Meanwhile, the output of this exploratory study would act as a substantial input for other research studies informed by critical agenda aiming at implementing changes in similar contexts. The findings would also inform researchers who work on designing AL profiles for different stakeholders involved in language assessments. It can guide them on the content of assessment literacy components required for teachers involved in EAP assessments. Academically, I believe the findings of this study will redress a gap in the assessment literature in Kuwait, as I have not come across any study that explored or addressed TAL in EPPs. Therefore I intend to communicate the finding of the current research to different communities. First, the study findings will be presented and discussed in the annual seminar conducted on the described context. Second, I intend to present the finding of the research in a well-established assessment conference to reach to the broader community of assessors. Finally, I intend to publish the current study in JEAP to contribute to building knowledge related to TAL.

## **5.5 Merits of the Study**

An essential aim of qualitative research is "to help people recover, and release themselves from the constraints of irrational, unproductive, unjust and unsatisfactory social structures that limit their self-development and self-determination" (Kemmis & McTaggart, 2000). Therefore, I believe that giving the

assessors the opportunity to reflect critically on their beliefs and their practices, in addition to sharing their concerns and dissatisfactions regarding observed assessment problems would act as enlightenment for further reformation in APs. Practically, it is worth mentioning that despite conducting the current research to understand the reason behind the observed problem; however, I intend to use the output of the current study as a starting point for further transformative research to enhance the assessment practices in EPPs. Professionally, teachers in similar contexts may benefit from the study findings, especially if the study output could be presented as a journal article or a presentation in academic conferences.

## **5.6 Limitations of the Study**

The study was limited to the EPP in the described context. It focused on exploring the assessors' assessment literacy only; no other teachers were involved in the current study. It is worth mentioning that the study was not questioning international standardized language assessment; it only focused on classroom-based assessments. I intentionally did not provide a detailed description of the programme or the participants to protect their identity. Kuwait is small country with very limited number of universities and EAP instructors could be identified easily if detailed information is provided about them. I added a general description of the EPP in the introduction chapter to provide the reader with general information about the nature of the programme, the learners, the teachers, and the adopted teaching and assessment practices. These general features, which I provided are common among EPPs and could not be used to identify the specific context of the study.

I believe that it would have enriched the study and its findings if the other teachers and learners were given a chance to take part in this study. Involving other teachers, who were not involved in APs, and learners to explore their views about the observed APs would have given more validation to the observed problem, but due to practical and administrative issues, I was not allowed to discuss or collect data from the learners. Moreover, involving the management's beliefs in the study about the observed APs would have added a different lens to the study, especially regarding imposed practices and justifications for adopting

them. This involvement would have created a channel of communication between assessors and management that is lacked in the current context. Finally, I believe that conducting the study as a study group involving all assessors together in each stage would have enriched the study and resulted in more fruitful discussion and argument; however, due to practicality, work nature, and ethical consideration I could not adopt this practice.

## Bibliography

- Abell, S. K., & Siegel, M. A. (2011). Assessment literacy: What science teachers need to know and be able to do. In D. Corrigan, J. Dillon, & R. Gunstone (Eds.), *The Professional knowledge base of science teaching* (vol. 12, pp. 205-221). Dordrecht: Springer.
- Adie, L. (2010). *Developing shared understandings of standards-based assessment: Online moderation practices across geographically diverse contexts*. PhD Thesis. Brisbane: Queensland University of Technology.
- Adie, L. (2013). The development of teacher assessment identity through participation in online moderation. *Assessment in Education: Principles, Policy & Practice*, 20(1), 91-106.
- Airasian, P. (2005). *Assessment in the classroom: A concise approach* (2nd ed.). Boston, MA: McGraw-Hill Company
- Airasian, P. W., & Russell, M. K. (2008). *Classroom assessment* (6th ed.). New York: McGraw-Hill.
- Alderson, J. C., Clapham, C., & Wall, D. (1996). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Alderson, P., & Morrow, V. (2004). *Ethics, social research and consulting young children*. Barking: Barnardo's.
- Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of pre-service and in-service teachers in educational measurement. *Asia-Pacific Journal of Teacher Education*, 39(2), 113-123.
- Allan, A. (1992). Developing of a scale to measure test-wiseness in EFL/ESL reading test-taker. *Language testing*, 9, 101-122.
- Allwright, D. (2005). Developing principles for practitioner research: The case of exploratory practice. *The Modern Language Journal*, 89, 353-366.
- Alnahdi, G. H. (2014). Educational Change in Saudi Arabia. *Journal of International Education Research*, 10 (1).
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Retrieved from <http://www.teststandards.org/>.
- American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Retrieved from <http://buros.org/standards-teacher-competence-educational-assessment-students>.
- Arksey, H., & Knight, P. (1999). *Interviewing for Social Scientists*. London: Sage.
- Assessment Reform Group. (2008). *Changing assessment practices: Process, principles and standards*. Retrieved from <http://www.aaia.org.uk/content/uploads/2010/06/ARIA-Changing-Assessment-Practice-Pamphlet-Final.pdf>.

- Association for Educational Assessment-Europe. (2012). *European framework of standards for educational assessment 1.0*. Rome: Edizioni Nuova Cultura.
- Atac, B. A. (2012). Foreign language teachers' attitude towards authentic assessment in language teaching. *Journal of Language & Linguistics Studies*, 8(2), 7-19.
- Atkinson, P., & Delamont, S. (2010). *Sage Qualitative Research Methods*. Sage: London.
- Australian Curriculum, Assessment and Certification Authorities. (1995). *Guidelines for assessment quality and equity*. Retrieved from <http://acaca.bos.nsw.edu.au/files/pdf/guidelines.pdf>.
- Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2005). Building and supporting a case for a test use. *Language Assessment Quarterly: An International Journal*, 2(1), 1-34.
- Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Bailey, K. M., & Brown, J. D. (1995). Language testing courses. What are they? In A. Cumming & R. Berwick (Eds.), *Validation in language testing* (pp. 236-256). Clevedon, UK: Multilingual Matters.
- Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389-407.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares, & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents, adolescence and education* (pp. 307-337). Greenwich: Information Age Publishing.
- Barnes, N., Fives, H., & Dacey, C. M. (2015). Teachers' beliefs about assessment. In H. Fives, & M. G. Gill (Eds.), *International handbook of research on teacher beliefs* (pp. 284-300). New York: Routledge.
- Bernstein, B. (1999). Vertical and horizontal discourse: An essay. *British Journal of Sociology of Education*, 20(2), 157-173.
- Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum, & F. Dochy. (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3-29). Boston, MA: Kluwer Academic.
- Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2002). *Assessment for learning: Putting it into practice*. London: Oxford University Press.
- Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2004). Working inside the black box: Assessment for Learning in the classroom. *Phi Delta Kappan*, 86.

- Black, P., & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80.
- Black, P., & William, D. (2009). Assessment and classroom learning. *Assessment in Education*, 5, 7-74.
- Boeije, H. (2010). *Analysis in Qualitative Research*. Sage Publications Ltd.
- Borg, S. (2003). "Research education" as an objective for teacher training. In B. Beaven & S. Borg (Eds.), *The role of research in teacher training* (pp. 41-48). Whitstable, Kent: IATEFL.
- Brindley, G. (2001). Language assessment and professional development. In C. Elder, A. Brown, E. Grove, K. Hall, N. Iwashita, T. Lumley, K. O'Loughlin (Eds.), *Experimenting with uncertainty. Essays in honour of Alan Davies* (pp. 126-136). Cambridge, UK: Cambridge University Press.
- Brookhart, S. M. (1999). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice*, 18(1), 5-13.
- Brookhart, S. M. (2009). Assessment and examinations. In L. J. Saha, & A. G. Dworkin (Eds.), *International handbook of research on teachers and teaching* (Vol. 21, part 2, pp. 723-738). New York: Springer.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12.
- Brown, G. T. L. (2008a). Assessment literacy training and teachers' conceptions of assessment. In C. Rubie-Davies, & C. Rawlinson (Eds.), *Challenging thinking about teaching and learning* (pp. 285-302). New York: Nova Science.
- Brown, G. T. L., Irving, E., & Keegan, P. (2014). *An introduction to educational assessment, measurement, and evaluation: Improving the quality of teacher-based assessment*. New Zealand: Dunmore Publishing Limited.
- Brown, J. D. (2005). *Testing in Language Programs: A Comprehensive Guide to English Language Assessment*. McGraw-Hill College.
- Brown, J. D., & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing*, 25, 349-384.
- Brown, J. D., & Coombe, C. (Eds.) (2016). *The Cambridge guide to research in language teaching and learning*. Cambridge: Cambridge University Press.
- Bruce, I. (2011). *Theory and concepts of English for Academic Purposes*. Basingstoke: Palgrave Macmillan.
- Bulter, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology*, 78, 210-216.
- Burns, R. B. (2000). *Introduction to research methods*. (4th edition). Australia: Longman.

- Cameron, C., Tate, B., Macnaughton, D., & Politano, C. (1998). *Recognition without Rewards*. Winnipeg, MB: Peguis Publishers.
- Camilli, G. (2006). Test fairness. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.) (pp. 221–256). Westport: American Council on Education & Praeger.
- Campbell, C., & Collins, V. L. (2007). Identifying essential topics in general and special education introductory assessment textbooks. *Educational Measurement: Issues and Practice*, 26(1), 9-18.
- Campbell, C., Murphy, J. A., & Holt, J. K. (2002). *Psychometric analysis of an assessment literacy instrument: applicability to pre-service teachers*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus.
- Canale, M. (1983a). 'From communicative competence to communicative language pedagogy. In C. Richards and R.W. Schmidt (eds), *Language and Communication* (p.2-27). London: Longman.
- Canale, M., & Swain, M. (1980). 'Theoretical bases of communicative approaches to second language teaching and testing.' *Applied Linguistics*, 1(1), 1-47.
- Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. London: Routledge.
- Celce-Murcia, M., Dornyei, Z., & Thurrell, S. (1995). 'Communicative competence: a pedagogically motivated model with content specifications.' *Issues in Applied Linguistics*, 2, 5-35.
- Chalhoub-Deville, M. (1997). 'Theoretical models, assessment frameworks and test construction.' *Language Testing*, 14(1), 3-22.
- Cheng, L. Rogers, T. W., & Wang, X. (2008). *Assessment and Evaluation in Higher Education*, 33(1), Routledge.
- Chiu, T. (2015). Personal statement in PhD applications: Gatekeepers' evaluative perspectives. *Journal of English for Academic Purposes*, 14, 63-73.
- Cohen, A. D. (1994). *Assessing Language Ability in the Classroom*. U.S.A.: Heinle & Heinle.
- Cohen, L., Manion, L. & Morrison, K. (2007). *Research Methods in Education* (6<sup>th</sup> edition). Routledge
- Coombe, C., Davidson, P., Gebril, A., Boraie, D., & Hidri, S. (2017). *Language Assessment in the Middle East and North Africa: Theory, Practice and Trends*. Dubai: TESOL Arabia.
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyloff (Eds.), *The Cambridge guide to second language assessment* (pp. 20–29). Cambridge: Cambridge University Press.

Cornish, L., & Jenkins, K. A. (2012). Encouraging teacher development through embedding reflective practice in assessment. *Asia-Pacific Journal of Teacher Education*, 40(2), 159-170.

Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching and Assessment*. Cambridge: Cambridge University Press. Available on-line: [http://www.coe.int/t/dg4/linguistic/Source/Framework\\_EN.pdf](http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf).

Cowie, B., Cooper, B., & Ussher, B. (2014). Developing an identity as a teacher assessor: Three student teacher case studies. *Assessment Matters*, 7, 64-89.

Cronbach, L. J. (1984). *Essentials of Psychological Testing*. 4th edition. New York: Harper and Row.

Crossman, J. (2007). The role of relationships and emotions in student perceptions of learning and assessment. *Higher Education Research and Development*, 26(3), 313-327.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. SAGE 1, Routledge. <https://doi.org/10.1080/02602930601122555>.

Cumming, A. (2009). Language assessment in education: tests, curricula, and teaching. *Annual Review of Applied Linguistics*, 29, 90-100.

Daniel, L. G., & King, D. A. (1998). Knowledge and use of testing and measurement literacy of elementary and secondary teachers. *The Journal of Educational Research*, 91(6), 331-344.

Davidson, F., & Lynch, B. (2002). *Testcraft: A Teacher's Guide to Writing and Using Language Test Specifications*. New Haven, CT: Yale University Press.

Davies, A. (2004) 'Introduction: language testing and the golden rule.' *Language Assessment Quarterly* 1, 2 & 3, 97-108.

Davies, A. (2008). Textbook trends in teaching language testing. *Language Testing*, 25(3), 327-347.

Dawson, C. (2007). *A practical guide to research methods: A user-friendly manual for mastering research techniques and projects*. Oxford: How to Books.

DeLuca, C. (2012). Preparing teachers for the age of accountability: Toward a framework for assessment education. *Action in Teacher Education*, 34(5/6), 576-91.

DeLuca, C., Chavez, T., & Cao, C. (2013). Establishing a foundation for valid teacher judgement on student learning: The role of pre-service assessment education. *Assessment in Education: Principles, Policy & Practice*, 20(1), 107-6.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice* 17, 419-438.



DeLuca, C., & Lam, C. Y. (2014). Preparing teachers for assessment within diverse classrooms: An analysis of teacher candidates' conceptualizations. *Teacher Education Quarterly*, 41(3), 3-24.

DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international standards and measures. *Educational Assessment and Accountability*, 28, 251-272. <https://doi.org/10.1007/s11092-015-9233-6>

DeLuca, C., & McEwen, L. (2007). *Evaluating assessment curriculum in teacher education programs: An evaluation process paper*. Paper presented at the annual Edward F. Kelly Evaluation Conference, April, in Ottawa, ON.

Denicolo, P., & Becker, L. (2012). *Developing Research Proposals*. London: SAGE.

Department for Education-United Kingdom. (2012). *Teachers' standards*. Retrieved from <http://www.education.gov.uk/schools/teachingandlearning/reviewofstandards/a00205581/teachers-standards1-sep-2012>.

Department of Education-Australia. (2012). *Australian professional standards for teachers*. Retrieved from <http://www.teacherstandards.aitsl.edu.au/OrganisationStandards/Organisation>.

Derrich, J., & Ecclestone, K. (2006). *Formative assessment in adult literacy, language and numeracy programmes: A literature review for the OECD*.

Djoub, Z. (2017). Assessment Literacy: Beyond Teacher Practice. In R. Al-Mahrooqi, C. Coombe, F. Al-Maamari & V. Thakur (Eds.), *Revising EFL Assessment: Critical Perspectives* (pp. 9-28). Switzerland: Springer.

Dornyei, Z. (2000). Motivation in action: Towards a process-oriented conceptualisation of student motivation. *British Journal of Educational Psychology*, 70, 519-538.

Dornyei, Z. (2002). *Questionnaire in second language research: Construction, administration, and processing*. London: Lawrence Erlbaum Associates.

Dornyei, Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. UK: Oxford University Press.

Earl, L. (2003). *Assessment as learning*. Thousand Oaks: Corwin.

Ecclestone, K. (2012). Assessment. In J. Arthur, & A. Peterson (Eds.), *The Routledge Companion to Education* (pp. 157- 168). London: Routledge.

Efeotor, V. (2017). Assessment of Learning, for Learning: A Critical Analysis of a Saudi Preparatory Year Assessment Model. In C. Coombe, P. Davidson, A. Gebril, D. Boraie, & S. Hidri (Eds.), *Language Assessment in the Middle East and North Africa: Theory, Practice and Future Trends* (pp.15-33). Dubai: TESOL Arabia.

Eisner, E. (1991). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*. New York. Macmillan.

- Ellis, R. (2006). Current Issues in the Teaching of Grammar: An SLA Perspective. *TESOL Quarterly*, 40(1), 83-107.
- Elmahjoub, A. (2017). The Effects of Evaluative Assessment on the Practices of Teaching and Learning. In C. Coombe, P. Davidson, A. Gebril, D. Boraie, & S. Hidri (Eds.), *Language Assessment in the Middle East and North Africa: Theory, Practice and Future Trends* (pp.113-123). Dubai: TESOL Arabia.
- Fan, Y. C., Wang, T. H., & Wang, K. H. (2011). A Web-based model for developing assessment literacy of secondary in-service teachers. *Computers & Education*, 57(2), 1727-1740.
- Firoozi, T., Razavipour, K., & Ahmadi, A. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, 9(2), 2-14. <https://doi.org/10.1186/s40468-019-0078-7>.
- Fives, H., & Buehl, M. M. (2012). Spring cleaning for the “messy” construct of teachers’ beliefs: What are they? Which have been examined? What can they tell us? In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook: Individual differences and cultural and contextual factors* (pp. 471-499). Washington, DC: American Psychological Association.
- Fleer, M. (2015). Developing an assessment pedagogy: The tensions and struggles in re-theorizing assessment from a cultural-historical perspective. *Assessment in Education: Principles, Policy & Practice*, 22(2), 224-246.
- Flowerdew, J., & Peacock, M. (2001). *Research Perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press.
- Forsberg, E., & Wermke, W. (2012). Knowledge sources and autonomy: German and Swedish teachers’ continuing professional development of assessment knowledge. *Professional Development in Education*, 38(5), 741-758.
- Francis, S., Hirsch, S., & Rowland, E. (1994). Improving school through study groups. *Journal of staff Development*, 13, 12-15.
- Frey, N., & Fisher, D. (2009). Using common formative assessments as a source of professional development in an urban American elementary school. *Teaching and Teacher Education*, 25(5), 674-680.
- Fulcher, G. (2010). *Practical Language Testing*. London: Hodder Education.
- Fulcher, G. (2012) Assessment Literacy for the Language Classroom, *Language Assessment Quarterly*, 9 (2), 113-132.
- Fulcher, G., & Davidson, F. (2007). Classroom assessment. In G. Fulcher & F. Davidson, *Language testing and assessment* (pp. 23–35). New York, NY: Routledge
- Furtak, E. M., Ruiz-Primo, M. A., Shemwell, J. T., Ayala, C. C., Brandon, P. R., Shavelson, R. J., & Yin, Y. (2008). On the fidelity of implementing embedded formative assessments and its relation to student learning. *Applied Measurement in Education*, 21(4), 360-389.

- Gilbert, N. (Ed.). (2006). *From postgraduate to social scientist: A guide to key skills*. London: Sage.
- Gillham, B. (2000). *Developing a Questionnaire*. London; Continuum
- Gillham, B. (2005). *Research Interviewing: The Range of Techniques*. Open University Press: Poland.
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179-195. <https://doi.org/10.15446/profile.v20n1.62089>.
- Given, L. M. (ed.) (2008). *The Sage Encyclopedia of Qualitative Research Method: Volumes 1 & 2*, Sage Publications, Newbury Park, California.
- Goldberg, L. R. (1992). The development of markers for the Big Five structure. *Psychological Assessment*, 4, 26-42.
- Goswami, D., & Stillman, P. (1987). *Reclaiming the classroom: Teacher research as an agency for change*. Upper Montclair, NJ: Boynton Cook.
- Gottheiner, D. M., & Siegel, M. A. (2012). Experienced middle school science teachers' assessment literacy: Investigating knowledge of students' conceptions in genetics and ways to shape instruction. *Journal of Science Teacher Education*, 23(5), 531-557.
- Graham, P. (2005). Classroom-based assessment: Changing knowledge and practice through preservice teacher education. *Teaching and Teacher Education*, 21(6), 607-621.
- Greenberg, J., & Walsh, K. (2012). *What teacher preparation programs teach about K-12 assessment: A review?* Washington, DC: National Council on Teacher Quality.
- Gu, P. Y. (2014). The unbearable lightness of the curriculum: What drives the assessment practices of a teacher of English as a Foreign Language in a Chinese secondary school. *Assessment in Education: Principles, Policy & Practice*, 21(3), 286-305.
- Guba, E. G., & Lincoln, Y.S. (1989), *Fourth Generation Evaluation*, Sage Publications, Newbury Park, California.
- Guba, E. G., & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (eds.), *Handbook of qualitative research* (pp. 105-117). London: Sage.
- Gunn, A. C., & Gilmore, A. (2014). Early childhood initial teacher education students' learning about assessment. *Assessment Matters*, 7, 24-38.
- Guskey, T. (2007). Multiple sources of evidence: An analysis of stakeholders' perceptions of various indicators of student learning. *Educational Measurement: Issues and Practice*, 26(1), 19-27.
- Hambleton, R. K. (1994). The rise and fall of criterion-referenced measurement? *Educational Measurement: Issues and Practice*, 13 (4), 21-26.

Hamp-Lyons, L. (1991). 'Scoring procedures for ESL contexts.' In Hamp-Lyons, L. (ed.) *Assessing Second Language Writing in Academic Contexts* (pp.241-276). Norwood, NJ: Ablex.

Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari, & J. Banerjee (Eds.), *Handbooks of applied linguistics [HAL]: Handbook of second language assessment* (pp. 413-427). Berlin, Germany/Boston, MA: De Gruyter Mouton.

Hasselgreen, A., Carlsen, C., & Helness, H. (2004). *European Survey of Language Testing and Assessment Needs. Part 1: General findings*. Gothenburg, Sweden: European Association for Language Testing and Assessment. Available from <http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf>

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.

Heaton, J. (1998). *Writing English language tests*. New York: Longman Inc.

Heigham, J., & Croker, R. A. (2009). *Qualitative Research in Applied Linguistics*. London. Palgrave Macmillan.

Hill, M. F., Cowie, B., Gilmore, A., & Smith, L. F. (2010). Preparing assessment-capable teachers: What should pre-service teachers know and be able to do? *Assessment Matters*, 2, 44-64.

Hill, M. F., Ell, F., Grudnoff, L., & Limbrick, L. (2014). Practise what you preach: Initial teacher education students learning about assessment. *Assessment Matters*, 7, 90-112.

Hipwell, P., & Klenowski, V. (2011). A case for addressing the literacy demands of assessment. *Journal of Language and Literacy*, 34 (2), 127–146.

Holbrook, A. (2008a). Response order effects. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (Vol. 2, pp. 754-756). Thousand Oaks, CA: SAGE.

Holbrook, A. (2008b). Acquiescence response bias. In P. J. Lavrakas (Ed.), *Encyclopedia of survey research methods* (Vol. 1, pp. 3-4). Thousand Oaks, CA: SAGE.

Hogan, R., & Smither, R. (2001). *Personality: Theory and applications*. Boulder, CO: Westview Press.

Holliday, A. (2001). *Doing and writing qualitative research*. London: Sage.

Howell, K. E. (2013). *Introduction to the Philosophy of Methodology*. London: Sage Publications.

Howley, M. D., Howley, A., Henning, J. E., Gilla, M. B., & Weade, G. (2013). Intersecting domains of assessment knowledge: School typologies based on interviews with secondary teachers. *Educational Assessment*, 18(1), 26-48.

Huang, L. (2018). A call for critical dialogue: EAP assessment from the Practitioner's perspective in Canada. *Journal of English for Academic Purposes*, 35, 70-84.

Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

Huhta, A., Hirvalä, T., & Banerjee, J. (2005). *European Survey of Language Testing and Assessment Needs. Part 2: Regional findings*. Gothenburg, Sweden: European Association for Language Testing and Assessment. Available from [http://users.jyu.fi/~huhta/ENLTA2/First\\_page.html](http://users.jyu.fi/~huhta/ENLTA2/First_page.html)

Hyland, K. (2002). Specificity revisited: how far should we go now? *Journal of English for Specific Purposes*, 21 (4), 385-395.

Hyland, K. (2006). *English for academic purposes: an advanced resource book*. London: Routledge.

Inbar-Lourie, O. (2008a). Constructing a language assessment knowledge base: A focus on language assessment courses, *Language Testing*, 25, 385-402.

Inbar-Lourie, O. (2008b). Language assessment culture. In E. Shohamy, & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.) (pp. 285-299). Springer Science + Business Media LLC.

Inbar-Lourie, O. (2013). Language assessment literacy. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 2923-2931). Oxford: Blackwell.

Interstate Teacher Assessment and Support Consortium (InTASC). (2011). *InTASC model core teaching standards: a resource for state dialogue*. Washington: Council of Chief State School Officers. Retrieved from [http://www.ccsso.org/Resources/Publications/InTASC\\_Model\\_Core\\_Teaching\\_Standards\\_A\\_Resource\\_for\\_State\\_Dialogue\\_DApril\\_2011\].html](http://www.ccsso.org/Resources/Publications/InTASC_Model_Core_Teaching_Standards_A_Resource_for_State_Dialogue_DApril_2011].html).

Ismael. D. A. (2017). Kurdish EFL Teachers' Assessment Literacy in Alternative Assessments and Its Influence on the Ethicality of Their Assessments. In R. Al-Mahrooqi, C. Coombe, F. Al-Maamari, & V. Thakur (Eds.), *Revising EFL Assessment: Critical Perspectives* (pp. 29-48). Switzerland: Springer.

Jarr, K.A. (2012). *Education practitioners' interpretation and use of assessment results*. University of Iowa. Retrieved from <http://ir.uiowa.edu/etd/3317>.

Jia, Y. (2009). Ethical Standards for language testing professionals: An introduction to five major codes. Shiken: *JALT Testing & Evaluation SIG Newsletter*, 13(2), 2-8.

Jin, Y. (2010). The place of language testing and assessment in the professional preparation of foreign language teachers in China. *Language Testing*, 27(4), 555-584.

Joint Advisory Committee. (1993). *Principles for fair student assessment practices for education in Canada*. Retrieved from [http://www2.education.ualberta.ca/educ/psych/crame/files/eng\\_prin.pdf](http://www2.education.ualberta.ca/educ/psych/crame/files/eng_prin.pdf).

Joint Committee for Standards on Educational Evaluation. (2015). *Classroom assessment standards: practices for PK-12 teachers*. Retrieved from [http://www.teach.purdue.edu/pcc/DOCS/Minutes/1215\\_Handouts/20130116/JCSSE\\_Assessment\\_Standards.pdf](http://www.teach.purdue.edu/pcc/DOCS/Minutes/1215_Handouts/20130116/JCSSE_Assessment_Standards.pdf).

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Westport: American Council on Education.

Kanjee, A., & Mthembu, J. (2015). Assessment literacy of foundation phase teachers: An exploratory study. *South African Journal of Childhood Education*, 5(1):142-168.

Kemmis, S., & McTaggart, R. (2000). Participatory action research. In N. K. Denzin, & Y. S. Lincoln (eds.), *Handbook of qualitative research*. (3rd edition). Thousand Oaks, CA: Sage.

Kershaw IV, I. (1993). *Ohio vocational education teachers' perceived use of student assessment information in educational decision-making*. Ohio State University.

Kiomrs, R., Abdolmehdi, R., & Naser, R. (2011). On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *English Language Testing*, 4(1), 156-161.

Klenowski, V. (2009). Australian Indigenous students: Addressing equity issues in assessment. *Teaching Education*, 20(1), 77–93.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.

Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual process theory of acquiescence. *Journal of Personality and Social Psychology*, 77, 379-386.

Kubiszyn, T., & Borich, G. (1996). *Educational testing and measurement: classroom application and practice* (5th ed.). New York: HarperCollins.

Kunnan, A. J. (2003). Test fairness. In M. Milanovic, & C. Weir (Eds.), *Select papers from the European Year of Languages Conference, Barcelona* (pp.27–48). Cambridge: Cambridge University Press.

Kunnan, A. J. (2004). Regarding language assessment. *Language Assessment Quarterly*, 1(1), 1.

Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197.

Lan, C., & Fan, S. (2019). Developing classroom-based language assessment literacy for in-service EFL teachers: The gaps. *Studies in Educational Evaluation*, 61, 112-122.

Lantolf, J. P. (2000). Second language learning as a mediated process. *Language Teaching*, (33)2, 79–96.

- Lantolf, J. P. (2009). Dynamic Assessment: the dialectic integration of instruction and assessment. *Language Teaching*, 42(3), 355-368.
- Lantolf, J., & Poehner, M. (2008). Dynamic assessment. In E. Shohamy, & N. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed.): Vol. 7. Language testing and assessment (pp. 273–284). New York: Springer Science + Business Media.
- Latif, M. W. (2017). Tertiary Level EFL Practitioners' Understanding of the Dynamic of Assessment Literacy: An Exploratory Study. In C. Coombe, P. Davidson, A. Gebril, D. Boraie, & S. Hidri (Eds.), *Language Assessment in the Middle East and North Africa: Theory, Practice and Future Trends* (pp.77-85). Dubai: TESOL Arabia.
- Latif, M. W. (2020). *Tertiary EFL Practitioners' Beliefs, Knowledge base and Practices of Assessment Literacy and Test Development - An Exploratory Study*. Unpublished Thesis. University of Exeter. United Kingdom.
- Lea, M., & Street, B. (2000). Student writing and staff feedback in higher education: an academic literacies approach. In M. Lea, & B. Stierer (Eds.), *Student writing in higher education* (pp. 32-46). Milton Keynes: Open University Press.
- Leahy, S., & William, D. (2012). From teachers to schools: Scaling up professional development for formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 49-72). London: SAGE.
- Leaph, K., Channy, M., & Chan, N. K. (2015). Cambodian ELT university practitioners' use of standardized tests for practice and assessment. *Language Education in Asia*, 6(1), 4-16.
- Lee, I. (2009). Ten mismatches between teachers' beliefs and written feedback practice. *ELT Journal*, 63(1), 13-22.
- Lee, I. (2011). Feedback revolution: what gets in the way? *ELT Journal*, 65(1).
- Leighton, J. P., Gokierto, R. J., Cor, M. K., & Heffernan, C. (2010). Teacher beliefs about the cognitive diagnostic information of classroom- versus large-scale tests: Implications for assessment literacy. *Assessment in Education*, 17, 7-21.
- Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice and change. *Language Assessment Quarterly*, 1(1), 19-41.
- Leung, C., & Scott, C. (2009). Formative assessment in language education policies: emerging lessons from Wales and Scotland. *Annual Review of Applied Linguistics* 29, 64-79.
- Levy-Vered, A., & Alhija, F. N.-A. (2015). Modelling beginning teachers' assessment literacy: The contribution of training, self-efficacy, and conceptions of assessment. *Educational Research and Evaluation*, 21(5-6), 378-406.
- Locastro. V. (2000). Teachers helping themselves: Classroom research and action research. *The language Teacher*, 18(2), 4-7.



- Looney, J. (2008). *Teaching, learning and assessment for adults: Improving foundation skills*. OECD.
- Lucas, R. E., & Baird, B. M. (2006). Global self-assessment. In M. Eid, & E. Diener (Eds.), *Handbook of multimethod measurement in psychology* (pp. 29- 42). Washington, DC: American Psychological Association.
- Lukin, L. E., Bandalos, D. L., Eckhout, T. J., & Mickelson, K. (2004). Facilitating the development of assessment literacy. *Educational Measurement: Issues and Practice*, 23(2), 26-32.
- Lynch, B., & Shaw, P. (2005). Portfolios, power and ethics. *TESOL Quarterly*, 39(2), 263-297.
- Maclellan, E. (2004). Initial knowledge states about assessment: Novice teachers' conceptualizations. *Teaching and Teacher Education*, 20(5), 523-535.
- Mahapatra, S. K. (2016). Tracking the development of teachers' language assessment literacy: A case study. In G. Pickering, & P. Gunashekar (Eds.), *Ensuring quality in English language teacher education* (pp. 106-114). New Delhi, India: British Council.
- Malone, M. E. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329-344.
- Manning, A. (2013). *EAP teacher assessment literacy* (Unpublished PhD Thesis). University of Leicester.
- Manning, A. (2014). *Investigating EAP assessment literacy*. Paper presented at the ALTE Conference, Paris.
- Mansory, M., & Meccawy, Z. (2017). Teachers' Role(s) in Summative Assessment in a Foundation Programme at a Saudi University's English Language Institute. In C. Coombe, P. Davidson, A. Gebril, D. Boraie, & S. Hidri (Eds.), *Language Assessment in the Middle East and North Africa: Theory, Practice and Future Trends* (pp.57-70). Dubai: TESOL Arabia.
- McAdams, D. P. (2000). *The person. An integrated introduction to personality psychology*. Fort Worth, TX: Harcourt.
- McKay, P. (2006). *Assessing Young Language Learners*. Cambridge: Cambridge University Press.
- McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation*, 7(8). Retrieved from <http://PAREonline.net/getvn.asp?v=7&n=8>, accessed November 21, 2009.
- McMillan, J. H. (2001). *Essential assessment concepts for teachers and administrators*. Thousand Oaks, California: Corwin Press, Inc.
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, 22(4), 34-43.



McMillan, J. H. (Ed.). (2007). *Formative classroom assessment: Theory into practice*. New York: Teachers College Press.

McNamara, T. F. (1996) *Measuring Second Language Performance*. London: Longman.

Mertler, C. A. (2003). *Pre-service versus in-service teachers' assessment literacy: does classroom experience make a difference?* In Annual meeting of the Mid-Western Educational Research Association, Columbus.

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, 12(2), 101–113. Doi: 10.1177/1365480209105575.

Mertler, C. A., & Campbell, C. (2005). *Measuring teachers' knowledge & application of classroom assessment concepts: development of the assessment literacy inventory*. In Annual meeting of the American Educational Research Association, Montreal.

Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003) *A Brief Introduction to Evidence-centered Design*. Research Report RR-03–16. Princeton, NJ: Educational Testing Service.

Mohan, B. A. (2011). Qualitative Research Methods in Second Language Assessment. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 752-767). New York: Routledge.

Montee, M., Bach, A., Donovan, A., & Thompson, L. (2013). LCTL teachers' assessment knowledge and practices: An exploratory study. *Journal of the National Council of Less Commonly Taught Languages*, 14, 1-32.

Moss, C. M., & Brookhart, S. M. (2009). *Advancing formative assessment in every classroom*. Alexandria, VA: ASCD.

Moss, P. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice* 22, 4, 13-25.

Nassaji, H., & Fotos, S. (2004). Current Developments in Research on the Teaching of Grammar. *Annual Review of Applied Linguistics*, 24,126-145.

National Board for Professional Teaching Standards. (2001). *What teachers should know and be able to do*. Retrieved from [http://www.nbpts.org/sites/default/files/documents/certificates/what\\_teachers\\_should\\_know.pdf](http://www.nbpts.org/sites/default/files/documents/certificates/what_teachers_should_know.pdf).

National Council for the Accreditation of Teacher Education. (2008). *Professional standards for the accreditation of teacher preparation institutions*. Retrieved from: <http://www.ncate.org/public/standards.asp>.

National Council on Measurement in Education (1995). *Code of professional responsibilities in educational measurement*. Retrieved from: [http://www.natd.org/Code\\_of\\_Professional\\_Responsibilities.html](http://www.natd.org/Code_of_Professional_Responsibilities.html).

- New Zealand Teachers Council. (2008). *Graduating teacher standards*. Retrieved from <http://www.teacherscouncil.co.nz/>.
- Nier, V. C., Donovan, A. E., & Malone, M. E. (2013). *Promoting assessment literacy for language instructors through an online course*. Retrieved from [http://www.cal.org/ecolt/pdfs/AB\\_Poster\\_10-21-2013\\_FINAL.pdf](http://www.cal.org/ecolt/pdfs/AB_Poster_10-21-2013_FINAL.pdf)
- Nieto, S. (2010). *The Light in Their Eyes: Creating Multicultural Learning Communities (10th Anniversary ed.)*. New York: Teachers College Press, Columbia University.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students* (6th ed.). Boston: Pearson Education, Inc.
- Oliver, D. G., Serovich, J. M., & Mason, T. L. (2005). Constraints and opportunities with interview transcription: Towards reflection in qualitative research. *Social Forces*, 84, 1273-1289. <https://doi.org/10.1353/sof.2006.0023>
- O'Sullivan, B. (Ed.). (2011). *Language Testing: Theories and Practices*. Basingstoke: Palgrave Macmillan.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Sage Publications, Inc.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Paulhus, D. L. (2002). Socially desirable responding: The evolution of a construct. In H. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 67-88). Hillsdale, NJ: Erlbaum.
- Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R.W. Robins, R.C. Fraley, & R.F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 224-239). New York: Guilford.
- Perpignan, H. (2001). *Teacher-written feedback to language learners: Promoting a dialogue for understanding*. Unpublished doctoral dissertation, Lancaster University, United Kingdom.
- Perpignan, H. (2003). Exploring the written feedback dialogue: A research, learning and teaching practice. *Language Teaching Research*, 7, 259-278.
- Perrett, G. (1990). The language testing interview: A reappraisal. In J. H. A. L. de Jong & D. K. Stevenson (Eds.), *Individualizing the assessment of language abilities* (pp. 255-238). Clevedon, England: Multilingual Matters.
- Pill, J., & Harding, L. (2013). Defining the language assessment literacy gap: Evidence from a parliamentary inquiry. *Language Testing*, 30(3), 381-402.
- Pitoniak, M. J., Young, J. W., Martiniello, M., King, T. C., Buteux, A., & Ginsburgh, M. (2009). *Guidelines for the assessment of English language learners*. Princeton, NJ: Educational Testing Service.

- Plake, B., Impara, J., & Fager, J. (1993). Assessment competencies of teachers: a national survey. *Educational Measurement: Issues and Practice*, 12(4), 10-39.
- Plake, B., & Jones, P. (2002). *Ensuring fair testing practices. The responsibilities of test sponsors, test developers, test administrators, and test takers in ensuring fair testing practices*. A paper presented at the February 2002 meeting of the Association of Test Publishers, Carlsbad, CA. Retrieved from: <http://www.testpublishers.org/assets/documents/Ensuring%20Fair%20Testing%20volume%204%20issue%201%20070202.pdf>
- Popham, W. J. (1994). The instructional consequences of criterion referenced clarity. *Educational Measurement: Issues and Practice*, 13, (4), 15-18.
- Popham, W. J. (1995). *Classroom assessment: What teachers need to know*. Boston: Allyn-Bacon.
- Popham, W. J. (2003). Test Better, Teach Better: *The Instructional Role of Assessment*. Alexandria, VA: Association for Supervisors and Curriculum Development.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4-11.
- Popham, W. J. (2011). Assessment literacy Overlooked: A Teacher Educator's Confession. *The Teacher Educator*, 46, 265-273.
- Popham, W. J. (2019). *Classroom assessment: what teachers need to know?* 9th ed. Boston: Pearson.
- Punch, K. F. (2005). *Introduction to Social Research*. 2<sup>nd</sup> Edition. Thousand Oaks, Calif: Sage.
- Purpura, J. E. (2016). Second and foreign language assessment. *Modern Language Journal*, 100 (Supplement 2016), 290-208. (Centennial Issue)
- Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator*, 36, 115-131.
- Rainey, I. (2000). Action research and the English as a foreign language practitioner: time to take stock. *Educational Action Research*, 8(1), 65-91.
- Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249-258.
- Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: a learning-focused interpretation. *International Journal of Applied Linguistics*, 16(2), 163-188.
- Reeves, D. B. (2007). *Ahead of the Curve: The Power of Assessment to Transform Teaching and Learning*. Bloomington, IN: Solution Tree Press.
- Richards, J. C., & Farrell, T. S. C. (2015). *Professional development for language teachers*. Cambridge: Cambridge University Press.

- Richards, J. C., & Renandya, W. A. (2002). Assessments. In J. C. Richards, & W. A. Renandya (Eds.), *Methodology in language testing: An anthology of current practice* (pp. 335-337). Cambridge: Cambridge University Press.
- Richards, J. C., & Schmidt, R. W. (2010). *Longman Dictionary of Language teaching and Applied Linguistics* (4<sup>th</sup> Ed.). London: Routledge.
- Richards, K. (2003). *Qualitative inquiry in TESOL*. London: Palgrave.
- Robins, R. W., & John, O. P. (1997). The quest for self-insight: Theory and research on accuracy and bias in self-perception. In R. Hogan, J. A. Johnson, & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 649-679). San Diego, CA: Academic Press.
- Rossman, G. B., & Rallis, S. F. (2003). *Learning in the field: An introduction to qualitative research* (2nd ed.). Thousand Oak, CJ: Sage Publications.
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivation: Classic definitions and new directions. *Contemporary Educational Psychology*, 25, 54-67.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Sato, M., Wei, C. R., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, 45(3), 669-700.
- Spada, N., & Lightbown, P. (2008). Form-Focused Instruction: Isolated or Integrated? *TESOL Quarterly*, 42(2), 181-207.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309-327.
- Schmitt, D., & Hamp-Lyons, L. (2015). The need for EAP knowledge in assessment. *Journal of English for Academic Purposes*, 18, 3-5.
- Schon, D. (1983). *The reflective practitioner: How professional think in action*. New York: Basic Books.
- Shepard, L. A. (2000). The role of assessment in a learning culture, *Educational Researcher*, 29(7), 4-14.
- Shepard, L. A. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 623-646). Westport, CT: Praeger.
- Shepard, L. (2007). Formative assessment: Caveat emptor. In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 279-303). Mahwah, NJ: Lawrence Erlbaum Associates.
- Shohamy, E. (1988). A proposed framework for testing the oral language of second/foreign language learners. *Studies in Second Language Acquisition*, 10(2), 165-179.

- Shohamy, E. (1996). Test impact revisited: washback effect over time. *Language Testing*, 13 (3), 298-317.
- Shohamy, E. (2001). *The power of tests*. Harlow, UK: Longman, Pearson.
- Shohamy, E. (2004). Reflections on Research Guidelines, Categories, and Responsibility. *TESOL Quarterly*, 38(4), 728-731.
- Shohamy, E., Or, I., & May, S. (2017). *Language Testing and Assessment. Encyclopedia of Language and Education* (3rd ed.). Auckland: Springer.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153–189.
- Silverman, D. (2001). *Interpreting qualitative data: Methods for analysing talk, text and interaction*. (Second edition). London: Thousand Oaks.
- Smith, K. (2011). Professional development of teachers a prerequisite for AFL to be successfully implemented in the classroom. *Studies in Educational Evaluation*, 37(1), 55-61.
- Stabler-Havener, M.L. (2018). Defining, conceptualizing, problematizing, and assessing language teacher assessment literacy. *Working Papers in Applied Linguistics & TESOL*, 18(1), 1-22.
- Stiggins, R. J. (1991). Assessment literacy. *The Phi Delta Kappan*, 72, 534-539.
- Stiggins, R. J. 1999. Are you assessment literate? *High School Magazine*, 6 (5), 20-23.
- Stiggins, R. J. (2001). *Student-involved classroom assessment*. Prentice Hall.
- Stiggins, R. J. (2002). Assessment crisis: the absence of assessment for learning. *Kappan Professional Journal*. Retrieved from <http://electronicportfolios.org/afl/Stiggins-AssessmentCrisis.pdf>.
- Stiggins, R. J. (2008). *Student-involved assessment FOR learning* (5thed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.
- Stoynoff, S. J. (2008). *Developments in English language assessment APEC strategic plan for English and other languages*. Mankato: APEC Human Resources Development Working Group, Minnesota State University.
- Talib, R., Kamsah, M. Z., Ghafar, M. N. A., Zakaria, M. A. Z. M., & Naim, H. A. (2013). *T-assess: Assessment literacy for Malaysian teachers*. Paper presented at the International Conference on Assessment for Higher Education across Domains and Skills, Kuala Lumpur.
- Taylor, L. (2013). Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing*, 30(3), 403-412.

- Thomas, D. R. (2003). *A general inductive approach for qualitative data analysis*. School of Population Health. University of Auckland.
- Threlfall, J. (2005). The formative use of assessment information in planning – the notion of contingent planning. *British Journal of Educational Studies*, 53(1), 54–65. doi: 10.1111/j.1467- 8527.2005.00283.x
- Tierney, R. D. (2010). *Insights into fairness in classroom assessment: Experienced English teachers share their practical wisdom* (PhD dissertation). University of Ottawa, Canada (UMI No 69109).
- Tierney, R. D. (2013). Fairness in classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 125-144). Thousand Oaks, CA: SAGE.
- Torrance, H., & Pryor, J. (1998). *Investigating formative assessment*. Milton Keynes, UK: Open University Press.
- Troudi, S., Coombe, C., & Al-Hamly, M. (2009). EFL teachers' views of English language assessment in higher education in the United Arab Emirates and Kuwait. *TESOL Quarterly*, 43(3), 546-555.
- Tsagari, D. (2004). Is there life beyond language testing? An introduction to alternative language assessment. *Centre for Research in Language Education*. CRILE Working Papers, 58. Retrieved from <http://www.ling.lans.ac.uk/groups/crile/docs/crile58tsagari.pdf>.
- Vazire, S., & Gosling, S. D. (2004). E-Perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology*, 87, 123-132.
- Vogt, K., & Tsagari, D. (2014). Assessment Literacy of Foreign Language Teachers: Findings of a European Study, *Language Assessment Quarterly*, 11:4, 374-402.
- Volante, L. (2006). Reducing bias in classroom assessment and evaluation. *Orbit*, 36(2), 34-36. Retrieved from: <https://ezpa.library.ualberta.ca/ezpAuthen.cgi?url=search.proquest.com/docview/213742716? Account id=144744>
- Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education* 30(3), 749-70.
- Wachter, B., & Maiworm. (2014). *English-taught programmes in European higher education: The state of play in 2014*. Bonn: Lemmens Accessed 10.04.15 [http://www.aca-secretariat.be/fileadmin/aca\\_docs/images/members/ACA-2015\\_English-Taught.pdf](http://www.aca-secretariat.be/fileadmin/aca_docs/images/members/ACA-2015_English-Taught.pdf).
- Wahyuni, D. (2012). The research design maze: understanding paradigms, cases, methods and methodologies. *Journal of Applied Management Accounting Research*, Vol. 10 (1), 69-80
- Walters, F. S. (2010). Cultivating assessment literacy: Standards evaluation through language-test specification reverse engineering. *Language Assessment Quarterly*, 7, 317-342.

- Weigle, S. (2002). *Assessing Writing*. Cambridge: Cambridge University Press.
- Weir, C. (1990). *Communicative Language Testing*. New York: Prentice Hall.
- Weir, C. (2005). *Language Testing and Validation: An Evidence-based Approach*. Basingstoke: Palgrave Macmillan.
- Westheimer, J. (2008). Learning among colleagues: Teacher community and the shared enterprise of education. In M. Cochran-Smith, S. Feiman-Nemser, & D. J. McIntyre (Eds.), *Handbook of research on teacher education: Enduring questions in changing contexts* (pp. 756-783). New York, NY: Routledge.
- William, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18-40). New York: Routledge.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3-14.
- Willis, J. (2010). Assessment for learning as a participatory pedagogy. *Assessment Matters*, 2, 65-84.
- Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualizing teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, 40(2), 241-256.
- Woolfolk, H. A., Hoy, W. K., & Davies, H. A. (2009). Teachers' self-efficacy beliefs. In K. Wentzel, & A. Wigfield (Eds.), *Handbook of motivation at school* (pp. 627-655). Mahwah, NJ: Lawrence Erlbaum.
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, 17(1), 59-75.
- Xu, Y., & Brown, G.T.L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149-158.
- Xu, Y., & Brown, G.T.L. (2017). University English teacher assessment literacy: A survey-test report from China. *Papers in Language Testing and Assessment*, 6(1), 133-158.
- Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 43(3), 493-513
- Yastıbaş, A. E., & Takkaç, M. (2018). Understanding language assessment literacy: Developing language assessment. *Journal of Language and Linguistic Studies*, 14(1), 178-193.
- Zhang, Z., & Burry-stock, J. A. (1997). *Assessment practices inventory: a multivariate analysis of teachers' perceived assessment competency*. In Annual meeting of the American Educational Research Association, Chicago.

## **Appendices**

### **Appendix 1: First Research Tool - Open-Ended Questionnaire**

#### **Open-Ended Questionnaire**

##### **Assessment Role and Assessment Preparation**

*The questionnaire aims to collect some information about your theoretical and practical assessment education, in addition to collecting some information about your current assessment role and practices.*

##### **Part One: Theoretical and Practical Assessment Education**

Please name your educational qualification (Bachelor's Holder, Masters' Holder, Doctorate/ PhD Holder) and any assessment certificates you have earned.

---

---

---

During your pre-service preparation, have you been exposed to any language assessment preparation? Please specify the name of the course, textbook, workshop, seminar, if any. If the answer is No, please specify the reason.

---

---

---

During your in-service preparation, have you been exposed to any assessment training before performing your assessment task? Please specify the training. If no, please state the reason for not being involved in any assessment training.

---

---

---

After you started your assessment tasks, have you been involved in any professional development related to assessments (on the job/of the job)? If yes, please specify. Furthermore, if no, please mention the reason.

---

---

---



## Part Two: Work Experience

Please describe your assessment role.

---

---

For how long have you been involved in this role?

---

---

How were you chosen for this assessment role?

---

---

Describe your assessment tasks in points.

---

---

---

---

---

---

---

## **Appendix 2: Second Research Tool - Open-Ended Checklist**

### **Checklist**

#### **Language Assessment Knowledge Base**

##### **Checklist Aim**

This open-ended checklist is meant to explore the participants' confidence with assessment knowledge base proposed by the (TALiP) framework not to evaluate their assessment literacy level. It is also meant to elicit information about the participants' views regarding the proposed base knowledge to understand and interpret their conceptions and beliefs about the proposed knowledge base and their need for this knowledge in their assessments contexts.

##### **Checklist Layout**

The checklist is divided into seven main sections. Each section addresses one assessment knowledge component as per the TALiP framework and involves sub-items that represents the sub-skills of each knowledge as per explored literature.

##### **General Instructions**

Participants are kindly requested to perform the following for each item:

1. Read the items under each section.
2. Choose only one answer for each item and provide at least one verbal justification for their choice:  
High Confidence (HC) Medium Confidence (MC) Low Confidence (LC).

3. For each item, the participants will choose, needed (N) or Not Needed (NN); however, they need to provide at least one reason for each choice verbally.
  - 3 The participants can ask for clarification on the provided items if they are not familiar with them or are not able to interpret the intended meaning.
  - 4 All provided verbal responses will be audio recorded.
  - 5 The interviewer can explain details about any item if requested by the participants.

## 1. Disciplinary Knowledge and Pedagogical Content Knowledge

*Knowledge about language content and how this content is learnt and taught. Any Assessment starts with specifying what to assess*

<b>Theoretical Knowledge</b>	<b>HC</b>	<b>MC</b>	<b>LC</b>	<b>Comment</b>	<b>N</b>	<b>NN</b>	<b>Reason for N/NN</b>
<p>1. Knowledge of second language learning theories:</p> <p><i>How learners learn a second language</i></p> <p><i>(e.g. Behaviorism, Cognitivism, Innatism, Constructivism, Information Processing, Interactionism, Multilingual Theories)</i></p>							
<p>2. Knowledge of second language teaching theories</p> <p><i>How teachers teach a second language</i></p> <p><i>(e.g. Audio Lingual Method, Community Language Teaching, Communicative Approach, Natural Approach, Content-based Instruction, Task-Based Language Teaching)</i></p>							
<p>3. Knowledge of learning strategies</p> <p><i>Kind of strategies that could be used with learners to assist in learning a second language</i></p> <ul style="list-style-type: none"> <li>▪ <i>Memory Strategies (Memorizing, Retrieving)</i></li> </ul>							

<ul style="list-style-type: none"> <li>▪ <i>Cognitive Strategies (Understanding, Verification, Deduction, Induction, Reasoning, Creation)</i></li> <li>▪ <i>Compensation Strategies (Translation, Avoidance, Gestures, Guessing)</i></li> <li>▪ <i>Metacognitive Strategies (Arranging, Regulating, self-directing, Evaluating)</i></li> <li>▪ <i>Social Strategies (Asking, starting conversations, cooperating)</i></li> <li>▪ <i>Affective Strategies (Lowering Anxiety)</i></li> </ul>							
<p>4.Knowledge of learning styles:</p> <p><i>Individual learning styles, which learners can use to assist in learning a language</i></p> <ul style="list-style-type: none"> <li>▪ <i>Physical Styles (Visual, Auditory, Verbal Kinaesthetic, Tactile)</i></li> <li>▪ <i>Communication Styles (extroverted, introverted, Solitary, Social)</i></li> <li>▪ <i>Thinking styles (global, analytic, reflective, intuitive, open)</i></li> </ul>							
<p>5.Knowledge of theories of language use (Language Models)</p> <p><i>Language theories specify what constitutes a language or the kind of constructs teachers teach and assess according to different language models</i></p> <ul style="list-style-type: none"> <li>▪ <i>Canale &amp; Swain Language</i> <ul style="list-style-type: none"> <li>✓ <i>Grammatical Competence (knowledge of grammar, lexis, morphology, syntax, semantics, and phonology)</i></li> <li>✓ <i>Sociolinguistic knowledge (the rules of discourse), and (the ability to overcome communicative difficulties).</i></li> <li>✓ <i>strategic Competence</i></li> </ul> </li> <li>▪ <i>Bachman's Model</i></li> </ul>							

✓ <i>language competence (grammatical Competence, textual Competence, Pragmatic Competence)</i>							
✓ <i>knowledge of the world (cultural and personal knowledge)</i>							
✓ <i>Strategic Competence (assessment component, planning component, execution component that helps using the language for communication)</i>							

## 2. Knowledge of Assessment Purpose, Content, and Methods

*Knowledge of \*Why you assess, \*What to assess, \*How to assess.*

### A. Theoretical Knowledge (Declarative Knowledge Used in Evaluating or Selecting Assessments)

Knowledge	HC	MC	LC	Comment	N	NN	Reason for N/NN
1.Knowledge of language testing and assessments history <i>Phases that language testing and assessment have gone through</i>							
2.Knowledge of language testing and assessment philosophies <i>Different philosophies behind assessment (for evaluation/for learning/for both)</i>							
3.Knowledge of the impact of adopting a specific testing or assessment philosophy							
4.Knowledge of major assessment qualities: <i>(Authenticity, Practicality, Interactiveness, Fairness, Ethics, Impact, Washback)</i>							
5.Knowledge of assessment types <i>(Traditional/Alternative, Norm-referenced/Criterion-referenced, Summative/Formative, Externally Mandated/Internally Mandated, Standardized/Classroom)</i>							

6.Knowledge of assessment methods:  (Tests, Portfolios, Performance Assessment, Self-Assessment, Peer Assessment, Checklists, Interviews, Questionnaires, Role-plays)							
7.Knowledge of major steps in language test development:  (Test Purpose, Construct Definition, Content Specification, Test Specification, Item writing)							
8. Knowledge of cognition taxonomies required for different test items or tasks  (Bloom's / Webb's)							

## B. Practical Knowledge (Skills, Procedural Knowledge Needed for Creating or Designing Assessments)

Skill	C	MC	NC	Reason	N	NN	Reason for N/NN
1.Identifying and stating language assessment purpose							
2.Writing goals and objectives of instructions and consequently their assessments							
3.Aligning curriculum objectives, instructions, and assessments							
4.Defining the language construct(s), which an assessment will elicit information about							



5.Utilizing alternative means of assessments to make decisions based on substantive information							
6.Constructing test specifications/blueprints to design parallel forms of a test							
7.Writing selected-response items such as multiple-choice, true/false, and matching							
8.Designing constructed–response items (for speaking and writing)							
9.Constructing well laid-out and perfectly legible items/tasks							
10.Providing clear, explicit, unambiguous instructions							
11.Providing examples to make candidates familiar with assessment techniques							
12.Including many items to enhance scoring reliability							
13.Writing test syllabuses to inform test users of test formats where applicable							
14.Designing assessments that are valid not only in terms of course content but also in course tasks							
15.Designing assessments that are reliable, authentic, fair, ethical, practical and interactive							
16.Incorporating technologies in assessing students							

### 3. Knowledge of Assessment Grading

<b>A. Theoretical Knowledge (Declarative Knowledge)</b>	<b>HC</b>	<b>MC</b>	<b>LC</b>	<b>Comment</b>	<b>N</b>	<b>NN</b>	<b>Reason for N/NN</b>
1. Knowledge of grading rational ( <i>Norm-, Criterion-Referenced</i> )							
2. Knowledge of methods that change scores into meaningful composites ( <i>Points, Percent, Grades, Proficiency Levels</i> ).							
3. Knowledge of necessary linear scaling, weighting scaling, weighting components, and precision of results.							
4. Knowledge of scoring techniques for objectively-marked testing							
5. Knowledge of the nature, purpose, and design of scoring rubrics for subjectively marked tests							
6. Knowledge of consistency and moderation principles for judgment-based assessments.							
<b>B. Practical Knowledge (Procedural Knowledge)</b>							

1. Constructing scoring schemes that quantify learners' performance into useful information for decisions about learners, classrooms, schools and districts							
2. Designing scoring keys with explicit instructions for closed-ended responses (right/wrong, checklist, multipoint methods including rubrics and rating scales).							
3. Designing scoring rubrics with explicit instructions for open-ended responses (holistic, analytic, primary trait scoring)							
4. Calculating consistency of scores/ standard error of measurement to assess scoring reliability for both closed-ended and open-ended responses							
5. Calculating scorer reliability through scorer reliability co-efficient to quantify the level of agreement among scorers							
6. Designing training workshops for scorers on acceptable approved responses							

#### 4. Knowledge of Feedback

<b>A. Theoretical Knowledge (Declarative Knowledge)</b>	<b>HC</b>	<b>MC</b>	<b>LC</b>	<b>Comment</b>	<b>N</b>	<b>NN</b>	<b>Reason for N/NN</b>
1.Knowledge of purposes and principles of feedback							
2.Knowledge of different types of feedback (descriptive, evaluative, supportive) with their respective functions							
3.Knowledge of the strengths and weaknesses of various kinds of feedback (task, process, metacognitive, and self-oriented) in facilitating student learning							
<b>B. Practical Knowledge (Procedural Knowledge)</b>							
1.Setting educational goals and constructing tasks that will move learners towards these goals							
2.Communicating to learners the criteria for success through different communication modes like telling, showing, and having learners discover							
3.Showing learners the gap between their current stage of development and the next/final stage of development							

4. Intervening in the learning process either by taking an “interventionist” or “interactionist” role							
5. Scaffolding their learning through different mediating exercises “graduated prompt”, “testing the limits”, and “mediated learning experience”.							
6. Using internet resources such as online tutorials and adapt contents to address students’ particular needs							
7. Using assessment methods to monitor, learning by collecting formal data (tests) and informal data (observing) of students’ language development							
8. Providing encouraging feedback that is meant to modify the learning process not just a score							
9. Monitoring, recording, and reporting student language development							
10. Evaluating how well learners have succeeded in achieving the final learning goals							
11. Improving instruction based on assessment results and feedback							

## 5. Knowledge of Peer & Self Assessments

<b>A. Theoretical Knowledge</b>	<b>HC</b>	<b>MC</b>	<b>NC</b>	<b>Comment</b>	<b>N</b>	<b>NN</b>	<b>Reason for N/NN</b>
1.Knowledge of benefits of involving students in assessment							
2.Knowledge of strategies of using self-and peer assessment in different assessment tasks							
<b>B. Practical Knowledge</b>							
1.Training learners to effectively participate in assessments							
2.Using self- and peer-assessments to promote learners' ability to notice their learning problems and gaps in their target learning goals							
3.Providing learners with criteria used in judging their work and make them use it to self or peer assess their work and their peers work							
4. Encouraging learners to produce their rating criteria in groups and use them to assess one another.							
5.Training learners to use portfolios, diaries, continuous cards, digital audios, video diaries, online blogs where learners are expected to collect samples of work and commentary to monitor their learning process and keep records of how their work improves and develops							

## 6. Knowledge of Assessment Interpretation & Communication

A. Theoretical Knowledge	HC	MC	NC	Comment	N	NN	Reason for N/NN
1. Interpreting data related to test design such as item difficulty and item discrimination							
2. Interpreting the concept of reliability in language assessment: <i>(Dependability, Item analysis, Reliability Threat)</i>							
3. Interpreting validity in language assessments <i>(Construct, Content, Criterion, Consequential Validity, and Validity as Argument)</i>							
4. Interpreting data from large-scale tests: <i>(namely means, modes, medians, bell curves and can calculate them)</i>							
5. Inferring students' strength and weakness based on collected data to communicate it to different stakeholders							
<b>B. Practical knowledge</b>							
1. Calculating item difficulty and item discrimination for close-ended items to obtain reliable scores							

2. Investigating facility and discrimination indices statistically							
3. Calculating test/item reliability, calculating inter/intra-rater reliability							
4. Investigating validity using statistical procedures							
5. Articulating interpretation of norm- and criterion-referenced assessment results to a variety of audience: student, school, directors							
6. Communicating interpretations using different reporting tools: Report cards, documents, criteria, guidance counselling.							
7. Using software such as Statistical Package to communicate results							
8. Participating in committee or school-wide discussions about reforms in assessment-related issues, curriculum, materials, grading policies, accountability policies, school evaluation and school evaluation.							
9. Having the skills to coach students to analyze their assessment results, track their learning, communicate about their learning, and plan next steps.							



## 7. Knowledge of Assessment Ethics

<b>A. Theoretical Knowledge</b>	<b>HC</b>	<b>MC</b>	<b>LC</b>	<b>Comment</b>	<b>N</b>	<b>NN</b>	<b>Reason for N/NN</b>
1.Knowledge of meaning and implications of critical language testing: power, ethics, and fairness							
2.Knowledge of how to work towards equity, non-discrimination, inclusion, and social justice							
3.Knowledge of codes and concepts of professional morality							
4.Knowledge legal and ethical responsibilities concerning the use, storage, and dissemination of assessment results							
<b>B. Practical Knowledge</b>							
1.Observing guidelines for ethics related to language assessment							
2.Treating all students or users of language assessment with respect							
3.Providing assessment practices that are fair and non-discriminatory through assuring distributive justice							
4.Avoiding construct-irrelevant controversial, inflammatory offensive or upsetting test material							

5. Adopting transparent language assessment practices by informing students of what, how, and why of assessments							
6. Involving other colleagues in assessment decision making and practices							
7. Implementing democratic practices by accepting and being open to constructive feedback from colleagues							
8. Using a range of assessment approaches and multiple measures that allows students to show their knowledge							
9. Defining a clear purpose for assessments, develop specifications, evaluate the content and conduct a field test examination							
10. Using tests' scores ethically through providing evidence of fairness via statistical procedures such as validity and reliability							
11. Clearly and honestly informing the inferences and decisions that derive from scores in assessments							
12. Using assessment results for feedback to influence language learning							
13. Evaluating the kind of washback that assessments can have on learning, teaching, curricula, and institutions							

14. Judging the consequences (intended or unintended) stemming from assessment in own context through examining "consequential validity."							
15. Implementing democratic language assessment practices by giving students the opportunity to share their voices about assessment							
16. Critiquing the impact and power, which standardized tests can have							

## **Appendix 3: Third Research Tool - Semi-Structured Interview**

### **Semi-Structured Interview**

**Comment on how the following assessment features are practised in your context and how far you are satisfied with each of them.**

1. Why do you assess students in your teaching context (Is it for evaluation/learning/ or both)?
2. What is the source of assessments in your teaching context (Are they internally or do you use externally mandated ones)?
3. What assessment methods are used to assess learners' language skills in your teaching context (tests/ portfolios/ interviews/ checklist/ projects/ presentation)
4. Who is involved in assessment design, is it a specifically chosen committee or all language instructors are involved in the process?
5. How are the following skills assessed in your teaching contexts?  
(Reading, Writing, Speaking, Listening)
6. Describe the assessment design process in your teaching context.
7. How are assessments administrated in your teaching context (drafting/ printing/ storage/ invigilating/ submissions/ collections)?
8. What kind of feedback do learners receive in your teaching context if any  
(Form, type, purpose)?
9. Do you apply peer or self-assessment principles and approaches in your teaching context? Why?
10. How are reading/ writing/speaking/listening assessments graded in your teaching context (keys/ rubrics/ scales)?
11. How are grading keys/rubrics/scales designed if any?
12. How are issues like assessment validity/reliability checked in your context? If available, how are they conducted?
13. How are assessments' results communicated to students?
14. How are assessments' fairness and ethical procedures practised in your teaching context if any?

**How far do you believe your context-adopted assessment practices reflect learners' language ability you are familiar with as a teacher? Why?**

## **Appendix 4: Fourth Research Tool - Open-Ended Report**

### **Open-Ended Report**

<b>Assessment Component</b>	<b>Current Practice</b>	<b>Participants' Belief</b>	<b>Solution to Conflicts</b>
Assessment Purpose			
Assessment Source			
Assessment Methods			
Assessment Decision Making			
Assessment Design Process			
Assessment of Language Skills			
Assessment Administration			
Assessment Grading			
Peer or Self-Assessment			
Assessment Feedback			
Assessment Result Communication			
Assessment Reliability			
Assessment Validity			
Assessment Fairness & Ethicality			

## **Appendix 5: Fifth Research Tool - Unstructured Interview**

### **Unstructured Interview Questions**

Comment on your experience with the different phases of the current research study.

Comment on the impact of such research experience on your assessment literacy. And how do you think your assessment literacy could be more developed and enhanced in the future?

Lets' imagine a different situation. If you were the assessment decision-maker in your context, what would be your future development and enhancement related to teacher assessment literacy?

## Appendix 6: Ethics Application Form



Ref (for office use only)

D1920-017

### COLLEGE OF SOCIAL SCIENCES AND INTERNATIONAL STUDIES

All staff and students within SSIS should use this form; those in Egenis, the Institute for Arab and Islamic Studies, Law, Politics, the Strategy & Security Institute, and Sociology, Philosophy, Anthropology should return it to [ssis-ethics@exeter.ac.uk](mailto:ssis-ethics@exeter.ac.uk). Staff and students in the Graduate School of Education should use [ssis-gseethics@exeter.ac.uk](mailto:ssis-gseethics@exeter.ac.uk).

Before completing this form please read the Guidance document which can be found at <http://intranet.exeter.ac.uk/socialsciences/ethics/>

Applicant details		
Name	Fatma Ismail Mohamed Ismail Mohamed	
Department	Graduate School of Education -EDD (TESOL)- Exeter Pathway (PT)	
UoE email address	Fii201@exeter.ac.uk	
Duration for which permission is required		
Start date:01/12/2019	End date:10/07/2022	Date submitted:29/08/2019
Students only		
Student number	620039569	
Programme of study	Doctor of Education (EdD) module	
Name of Supervisor(s) or Dissertation Tutor	Dr. Philip Durrant	
Have you attended any ethics training that is available to students?	Yes, I have taken part in ethics training at the University of Exeter 1- During the master's program at University of Exeter (Master in Education (TESOL) 2- During the EdD programme (TESOL) – (PT) - Exeter Pathway. During the Taught Module Phase EEDD042 16/07/2017	
Certification for all submissions		
I hereby certify that I will abide by the details given in this application and that I undertake in my research to respect the dignity and privacy of those participating in this research. I confirm that if my research should change significantly I will seek advice, request approval of an amendment or complete a new ethics proposal. Any document translations used have been provided by a competent person with no significant changes to the original meaning. I hereby acknowledge that I have read the British Educational Research Association's Ethical Guidelines (Fourth Edition, 2018) and will abide by all guidelines in the handbook. <a href="https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018">https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018</a> <b>Fatma Ismail Mohamed Ismail Mohamed</b> Double click this box to confirm certification ☐ <i>Submission of this ethics proposal form confirms your acceptance of the above.</i>		

**TITLE OF YOUR PROJECT**

*Exploring EFL Teachers'/Assessors' Assessment Literacy in an English Preparatory Programme in Kuwait*

**ETHICAL REVIEW BY AN EXTERNAL COMMITTEE**

*No, my research is not funded by, or doesn't use data from, either the NHS or Ministry of Defence.*

**MENTAL CAPACITY ACT 2005**

No, my project does not involve participants aged 16 or over who are unable to give informed consent (e.g. people with learning disabilities)

**SYNOPSIS OF THE RESEARCH PROJECT**

*Maximum of 750 words.*

The research aims to understand, re-conceptualize, and possibly develop EFL teachers'/ assessors' assessment literacy in practice. This empirical research, which is conducted with eight EFL teachers/ assessors in an English preparatory programme in Kuwait used a conceptual framework of teacher assessment literacy in practice (TALiP) proposed by Xu & Brown (2016) as an operational model for researching and exploring TALiP in the described context.

To achieve the research aim, the study seeks answers to the following research questions that address the components of the adopted conceptual framework of assessments in practice:

- 1.A. How are EFL teachers/assessors prepared for assessment tasks?
- 1.B. How are EFL teachers/assessors involved in assessment tasks in their teaching context?
- 2.A. How confident are EFL teachers/assessors about assessment knowledge base suggested by the TALiP framework?
- 2.B. What are EFL teachers'/assessors' views about assessment knowledge base suggested by the TALiP framework?
- 3.A. How are learners' EFL skills assessed in the participants' teaching context?
- 3.B. What are EFL teachers'/assessors' views about their context-adopted assessment practices that are used in assessing learners' EFL skills?
4. How can the participants negotiate between their views about assessments and their context-adopted assessment philosophies and practices if they don't agree?
5. How far have the participants' assessment literacies been developed as result of taking part in the research?

The research questions are designed to fulfil the following research objectives:

- a. Explore and understand the assessors' formal/informal assessment preparation whether during pre-service or in-service phase in addition to their current involvement in assessment practices in their teaching context.
- b. Expose the participants to assessment theoretical knowledge base as per explored literature, which they might be familiar with or not. If they are not, such exposure will add to their theoretical knowledge and will open up topics for them to reflect on and to consider in their assessment practices.
- c. Elicit their beliefs about what is needed or appropriate from the presented assessment knowledge base as per their conceptions. This is an important step because they will justify reasons for their views because any knowledge base consists of "core knowledge "that is applicable to all contexts/subjects/disciplines and "subject/context dependent knowledge", which cannot be elicited except in its contexts and from its actors. (what is needed and what is not needed and why)
- d. Elicit information on how learners' language skills are assessed in their contexts.



- e. Elicit their views about their context-bond assessment practices and how assessment could be best practiced in their teaching contexts as per their theoretical and practical experience, this is one important step in the research because it will reveal the participants' individual views.
- f. Give the participants the opportunity to reflect on their context-bond assessment practices and compare them to their own beliefs, a stand point that teachers/assessors need to experience.
- g. Assist in participants' decision making through putting them in problem solving situations where they engage in a collaborative professional dialogue with colleagues to come up with solutions to conflicts that arouse between their views and contextual assessment practices.
- h. Develop the participants' assessment literacy by making them the main drivers of their own development to be able to make use of the research experience in their class-based assessment through asking them to reflect on their participation and utter areas of development they experienced as a result of taking part in the research

#### INTERNATIONAL RESEARCH

Given the fact that this research is conducted with the participants about their own assessment literacy and development, I do not intend to make applications to the Research Ethics Committee in the described context, as they don't require permission for conducting research with teachers. Permission is only required if the research involves collecting data from students.

#### RESEARCH METHODS

In order to address the research aims and questions, different research tools will be used to collect self-reported qualitative data from eight participants in the described context about their assessment literacy in practice. The collected qualitative data will be subjected to interpretive analysis, which is expected to provide answers to the research questions, and therefore address the research aims and goals.

To answer the first research question, the study will use a self-administrated questionnaire with open ended questions sent by email to the eight participants to collect qualitative data about their assessment formal/informal preparation and their assessment roles in their teaching context; and therefore provide answers to the first research question.

To find answers to the second research question, the researcher will use a face to face audio-taped structured-interview with an interview schedule. The interview schedule is structured in the form of an open-ended checklist to elicit verbal qualitative data about the participants' confidence with the assessment base knowledge suggested by the TALiP framework and their beliefs about the component of the suggested base knowledge. The checklist is designed in a way that reflects the component of the first level of TALiP framework (Assessment Base knowledge). It is divided into seven main parts. Each part addresses one assessment knowledge component as per the TALiP framework as follows:

1. Disciplinary knowledge and Pedagogical Content Knowledge
2. Knowledge of Assessment Purpose, Content, and Methods
3. Knowledge of Assessment Grading
4. Knowledge of Feedback
5. Knowledge of Peer & Self Assessments
6. Knowledge of Assessment Interpretation & Communication
7. Knowledge of Assessment Ethics

Each part of the checklist involves sub-items that represents the sub-skills/-knowledge of each knowledge component and the participants are required to choose a confidence level (high, medium, low) in addition to verbally justifying or commenting on their choice. They're also required to comment on the relevance

of the items to their assessment contexts by choosing (needed or not needed); in addition to justifying their choices.

To elicit answers to the third research question, the researcher will conduct an audio-taped face to face semi-structured interview with an interview guide. The guide will include questions that elicit information about the participants' context-adopted assessment practices that are used to assess learners' language skills and their views about such practices. The questions will try to elicit information about the adopted assessment philosophy, purpose, methods, techniques, grading, administration, feedback, and communication and the participants' views about such practices.

To elicit answers to the fourth research questions, the researcher will use an open-ended report, which will be sent by email to the participants to fill out. The report is designed based on the collected data from the previous semi-structured interview. The data will be written in the form of a report for each participant (eight reports). Each report will be divided into themes that reflect the components of assessment explored in the previous interview (philosophy, purpose, method, design, content, grading, feedback, communication, self and peer assessment and ethic). In front of each theme, the researcher will add three columns: how the theme is practiced in the participants' context as per the participants' analyzed responses, the participant's view about such practice as provided in the semi-structured interview, and an empty column, where the participants are required to add their critical reflective ideas on how to compromise between their adopted practices and their beliefs if conflicts or discrepancies are detected. Through the reports, the participants will be put in a problem-solving situation, where they will be required to critically reflect on assessment problems they face in their contexts; they are expected to come up with solutions to conflicts they might face in real life. The report is expected to provide information on how the participants believe the assessment situation should be practiced in a better way. This is the most important part of the study because it involves understanding, enlightenment, and development that could happen to both the researcher and the participants, which is the main aim of the research study. It will be administrated this way because the researcher wants to provide the participants with the chance to read through the report, think about their answers and faced challenges, and then come up with possible solutions or ideas.

To elicit answers to the fifth research questions, the researcher uses an audio-recorded face to face unstructured interview. The aim of the interview is to elicit information on how the participants' AL would be developed as a result of going through the TALiP levels and taking part in the current exploratory study. The participants will be asked to talk freely about their experience with the current study and its impact on them. They will be given an imaginative scenario, where they imagine themselves the only assessment decision makers in their teaching contexts, and they will be asked to elaborate and talk about that new role based on their new developed literacy that they acquired as a result of taking part in the study and going through the different research levels. The nature of the used questions will invite description, explanation, reflection, and interpretation of the research experience (Patton, 2002: 348-351); the researcher will try indirectly to elicit the participants' experience, opinions, and feelings about the current research practices to reach a noticing phase, at which the participants can recognize and realize the kind of development that might happen to them.

## **PARTICIPANTS**

Eight experienced assessors, who are involved in assessment decision-making and design in the described context will be chosen to take part in the study, since it is believed that their shared knowledge will be relevant to the topic under investigation. The participants are considered key informants because they are members of the assessment committee in the described context. They are performing different assessment tasks; they are responsible for assessment design, method and content selection, rubric and answer key design, distribution of grading tasks to other teachers, administrating assessments, result communication and interpretation. They play key role in assessment decision making, which affects

inferences made about students' performance and consequently impacts students' future and continuation of studies in the described context. They are typical homogenous sample (Dornyei, 2007) that share common core experience and possess important experience that is relevant to the current study. Therefore, it is believed that exploring their knowledge, conceptual beliefs, views about assessment in general and their context-bond adopted practices, in addition to their critical views about the whole assessment process and their own assessment literacy would provide key transformative knowledge, about the assessment process in the studied context and how could it be better enhanced, which is one goal of the current study. The shared experience of the participants would explain the observed problem of assessment in the studied context. It would allow both the researcher and the participants to negotiate meaning in relation to assessment philosophies and practices and engage them in a professional dialogue that is meant for mutual conceptual development of assessment and assessors' literacies. Finally, it would open the gate for future assessment developmental steps as a result of reconceptualizing of assessment concepts that would possibly happen as a result of taking part in the different research procedures and stages described ahead.

#### **THE VOLUNTARY NATURE OF PARTICIPATION**

Once approval is received on the ethic application form from the university, all participants will be contacted to take their initial approval to take part in this study. Upon receiving verbal approvals from the participants, information sheet and consent forms will be prepared for each participant. Before conducting any data collecting procedure, a brief meeting will be conducted with each participant to explain the five steps of data collection procedures and the purpose behind each step. Each participant will be asked to decide on convenient time and place for conducting the interviews. Accordingly, an interview appointment schedule will be created. The appointment schedule for the three interviews will be confirmed by each participant according to their preferred time and location. During the same meeting, the researcher will collect the participants' email address, through which he/she prefers to be contacted for sending soft copies of the open-ended questionnaire and the open-ended report. Each participant will be provided with hard copies/templates, and the researcher will explain briefly what will be required for each document. The participants' feedback will be collected regarding clarity, layout items, questions included in both documents. The documents will be adjusted as per received back. They will be informed that the two document would be sent to the provided email and they would reply back with the filled document as an attachment within one week for the first one and two weeks for the second one. For the interviews, a reminder will be sent to each participant two days ahead of each meeting. For any inconvenience, the time would be rescheduled as per convenient time.

During the meeting, the researcher and the participants will agree on audio- recording the interviews. It will be agreed that each interview would last for around an hour and would be conducted under the same conditions. English would be used as the language for the interview because this is the common language between the researcher and the participants as all participants are competent speakers of the English language. Finally, the participants will be informed that they have the right to refuse to answer any question or withdraw from the study without offering any explanation.

#### **SPECIAL ARRANGEMENTS**

The research does not require special arrangement apart from arranging for the data collection schedule and location with the participants.

#### **THE INFORMED NATURE OF PARTICIPATION**

All participants are provided with an information sheet, which outlines the project aim, purpose, participants' roles, and participants' rights, in addition to protection steps adopted by the researcher to protect their identity, shared data, and safety. The sheet also includes the researchers', research supervisor's and ethic chairman's contact details for the participants to contact if they feel unhappy about any research procedures they might experience before, during, or after data collection procedures. The document will inform the participants that their names and the programme title will be kept anonymous

and replaced by pseudonyms. All participants will be informed that the data will be collected and analysed specifically for the purpose of the research. They will be informed that none of the output (the dissertation, article, report, conference, seminar, and presentation) would provide information that would identify their names or their teaching contexts. The information sheet will be given to each participant before data collection procedures, along with a consent form with the title and their detailed rights written in points to be initialled, dated and signed by them.

#### **ASSESSMENT OF POSSIBLE HARM**

I don't believe that taking part in the research would have any foreseeable risks on the participants. The provided data would be analysed specifically for the research. None of the output (the dissertation, article, report, conference, seminar, presentation) would provide information that would identify their names or names of their working context. For safety and security issues, all hard-copy transcripts will be stored in a locked filing cabinet, and the audio data will be deleted immediately from the digital recorder after being stored on a password-protected stationary device.

#### **DATA PROTECTION AND STORAGE**

Hard copy transcripts will be stored in a locked filing cabinet and audio data will be downloaded from recording devices at the earliest possible opportunity, and then deleted immediately from those devices. Furthermore, all audio data and soft copies of the transcripts will be stored and password protected on a password protected stationary device (PC) and NOT on a memory stick. Another copy will be saved on the university cloud storage.

Data will be kept until end of project and result announcement, then and all hard and soft copies of data will be destroyed by deleting all soft files and shredding all hard copies transcripts.

#### **DECLARATION OF INTERESTS**

This research is self-funded and is conducted by a practitioner for the sake of own and colleagues' professional understanding and development.

#### **USER ENGAGEMENT AND FEEDBACK**

The researcher will collect feedback from the participants through the following:

A "member check" technique by providing the participants with the research report based on the transcription of their interviews to indicate their agreement or disagreement with the collected data.

Their feedback on the report will be taken into consideration and report will be adjusted accordingly before data analysis. Record of such interactions, feedback and adjustment will be reported as part of the research journal.

#### **INFORMATION SHEET**

Information Sheet is attached to the report

#### **CONSENT FORM**

Consent Form is attached to the report

#### **SUBMISSION PROCEDURE**

Staff and students should follow the procedure below.

**Post Graduate Taught Students (Graduate School of Education):** Please submit your completed application to your first supervisor.

**All other students** should discuss their application with their supervisor(s) / dissertation tutor / tutor and gain their approval prior to submission. Students should submit evidence of approval with their application, e.g. a copy of the supervisors email approval.

**All staff** should submit their application to the appropriate email address below.

This application form and examples of your consent form, information sheet and translations of any documents which are not written in English should be submitted by email to the SSIS Ethics Secretary via one of the following email addresses:

[ssis-ethics@exeter.ac.uk](mailto:ssis-ethics@exeter.ac.uk) This email should be used by staff and students in Egenis, the Institute for Arab and Islamic Studies, Law, Politics, the Strategy & Security Institute, and Sociology, Philosophy, Anthropology.

[ssis-gseethics@exeter.ac.uk](mailto:ssis-gseethics@exeter.ac.uk) This email should be used by staff and students in the Graduate School of Education.

Please note that applicants will be required to submit a new application if ethics approval has not been granted within 1 year of first submission.

## **Appendix 7: Certificate of Ethical Approval**



**GRADUATE SCHOOL OF EDUCATION**

St Luke's Campus  
Heavitree Road  
Exeter UK EX1 2LU

<http://socialsciences.exeter.ac.uk/education/>

### **CERTIFICATE OF ETHICAL APPROVAL**

Title of Project:

Exploring EFL Teachers'/Assessors' Assessment Literacy in an English Preparatory Programme in Kuwait

Researcher(s) name: Fatma Ismail Mohamed

Supervisor(s): Philip Durrant

This project has been approved for the period

From: 01/12/2019

To: 10/07/2022

Ethics Committee approval reference: D1920-017

Signature:  Date: 12/10/2019  
(Professor Dongbo Zhang, Graduate School of Education Ethics Officer)



## **Appendix 8: Participant Information Sheet**



### **Participant Information Sheet**

(Participant:.....)

**Title of Project: Exploring EFL Teachers'/Assessors' Assessment Literacy in an English Preparatory Programme in Kuwait**

**Researcher name: Fatma Ismail Mohamed Ismail Mohamed**

#### **Invitation and brief summary**

*This exploratory study is conducted with EFL teachers/assessors to explore, understand and reconceptualise their assessment philosophies, views, and practices. Taking part in the study will allow you and the researcher to have a better understanding of your assessment philosophies and practices and go through reflective practice regarding those views and practices to be able to enhance and develop them. Please take time to read the provided information on the sheet and feel free to ask any questions before we start.*

#### **Purpose of the research**

*The research aims to understand, reconceptualise, and possibly develop EFL teachers'/ assessors' assessment literacy in practice. This research is self-funded and is conducted by a practitioner for the sake of own and colleagues' professional understanding and development. Taking part in the study will get you engaged in a professional dialogue that will provide you with the opportunity to reflect on your practices, develop your understanding, and get your voice heard by other communities. You will contribute to EFL assessment knowledge building through sharing recommendations, suggestions, and views with other practitioners. Sharing this information will provide a vital understanding of the nature of assessments in English preparatory programs to be shared with broader communities for mutual benefit and reflection. Very few studies were conducted on assessors' perceptions and beliefs as a critical component in framing the conceptual understanding of their assessment literacy. Therefore, this study will fill a gap in the literature of assessment literacy and will include assessors' voice and role in conceptualising their literacies and practices. Sharing the finding of the research in the form of an understanding and consensus regarding EFL assessment will act as a sort of professional development document for pre-service and novice in-service EFL practitioners once the study is published and presented for assessment communities in academic conferences.*

**Why have I been approached?**

*I approached you because I believe you are a key informant for the study. Taking into consideration your educational qualification, assessment background, experience with EFL contexts, and EFL assessment role, it is believed that the information you will provide will be a valuable source of understanding of the topic under investigation. Besides, I believe this is a valuable opportunity for self-development through being engaged in a professional dialogue about a practice you are engaged in. Working with you in the same context and on the same practice encouraged me to approach you to take part in this exploratory practice for mutual benefit, understanding and development, which will be reflected on our assessment practice and teaching contexts.*

**What would taking part involve?**

*You are expected to take part in three interviews, each will last for around an hour, in addition to filling a soft questionnaire and a report. The interviews will be conducted in English and will be audio-taped for transcription and analysis purposes. You will be approached throughout the interview using a pseudonym to protect your identity. You will not refer to the name of your teaching context throughout the interviews. During the research procedures, you will share your information about your assessment knowledge, philosophies, and beliefs, your views about your context-adopted assessment practices, how you compromise between both if any conflicts occur, and finally recommendations on better EFL assessment practices based on critical reflection practise you are engaged in throughout the research procedures. You are expected to respond to questions in an open-ended questionnaire, an open-ended report, a structured interview, a semi-structured interview, and an open-ended interview. You can ask for clarification or paraphrasing questions during any of the pre-mentioned research procedures.*

**What are the possible benefits of taking part?**

*Taking part in the study will get you engaged in a professional dialogue that will provide you with the opportunity to reflect on your practices, develop your understanding, get your voice heard by other communities. You will contribute to EFL Assessment knowledge building through sharing recommendations, suggestions, and views with other practitioners.*

**What are the possible disadvantages and risks of taking part?**

*I do not believe that taking part in the research would have any foreseeable risks on you. The provided data would be analysed specifically for the research. None of the output (the dissertation, article, report, conference, seminar, presentation) would provide information that would identify your name or name of your working context. For safety and security issues, all hard-copy transcripts will be stored in a locked filing cabinet, and the audio data will be deleted immediately from the digital recorder after being stored on a password-protected stationary device.*



**What will happen if I do not want to carry on with the study?**

*You can stop taking part at any time without having to give a reason. You need to inform me at least one day before the interview scheduled time if you decided to withdraw before conducting the interview. If you decided to withdraw after conducting the interview, you need to contact me through the below-specified contact details. I will provide you with a written declaration that your provided data will not be used for analysis or referred to in the research findings, and that your audio-taped interview or any provided soft document will be destroyed.*

**How will my information be kept confidential?**

The University of Exeter processes personal data to research public interest. The University will endeavour to be transparent about its processing of your data, and this information sheet should provide a clear explanation of this. If you do have any queries about the University's processing of your data that cannot be resolved by the research team, further information may be obtained from the University's Data Protection Officer by emailing [dataprotection@exeter.ac.uk](mailto:dataprotection@exeter.ac.uk) or at [www.exeter.ac.uk/data-protection](http://www.exeter.ac.uk/data-protection).

*Hard copy transcripts and soft documents will be stored in a locked filing cabinet, and audio data will be downloaded from recording devices at the earliest possible opportunity, and then deleted immediately from those devices. Furthermore, all audio data and transcripts will be stored, and password protected on a password protected stationary device (PC) and NOT on a memory stick. Data will be kept until the end of the project and result announcement, then and all hard and soft copies of data will be destroyed by deleting all soft files and shredding all hard copies transcripts.*

**Will I receive any payment for taking part?**

*Your participation is voluntary. You will not receive any payment for taking part in the project.*

**What will happen to the results of this study?**

*I intend to disseminate the results of the study in an academic publication and conference.*

**Who is organising and funding this study?**

*This research is self-funded and is conducted by a practitioner for the sake of own and colleagues' professional understanding and development.*

**Who has reviewed this study?**

*This project was reviewed by the Research Ethics Committee at the University of Exeter and the first supervisor of the research project.*

**Further information and contact details**

*Below are details of my contacts, the research supervisor, and Ethics Officer Chair in case you are not happy with any aspect of the project and wish to complain*

1- Researcher: Fatma Ismail Mohamed Ismail

Tel: 965 979172015

Email: [fii201@exeter.ac.uk](mailto:fii201@exeter.ac.uk)

2- Supervisor: Dr. Philp Durrant

Email: [P.L.Durrant@exeter.ac.uk](mailto:P.L.Durrant@exeter.ac.uk)

3- Research Ethics and Governance Manager: Gail Seymour

Email: [g.m.seymour@exeter.ac.uk](mailto:g.m.seymour@exeter.ac.uk),

Tel: 01392 726621

Thank you for your interest in this project

## Appendix 9: Consent Form



Participant Identification Number: (.....)

### CONSENT FORM

Title of Project: **Exploring EFL Teachers'/Assessors' Assessment Literacy in an English Preparatory Programme in Kuwait**

Name of Researcher: **Fatma Ismail Mohamed Ismail Mohamed**

Please initial box

1. I confirm that I have read the information sheet dated..... Version .....for the above project. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason and without my legal rights being affected. ☐
3. I understand that relevant sections of the data collected during the study, maybe looked at by members of the research team, individuals from the University of Exeter, where it is relevant to my taking part in this research. ☐  
I permit these individuals to have access to my records.
4. I understand that taking part in the research involves filling questionnaires and reports, in addition to participating in interviews, which will be audiotaped.  
The data provided will be subjected to transcription after which I will receive a report summarizing the collected data on which I can give feedback that involves agreement or disagreement to be used for research data analysis and findings. ☐
5. I understand that the research will last for an approximate period of four months that will involve several face-to-face meetings and electronic mail correspondence.  
The researcher will be using a pseudonym to address me during the research procedures and data collection phase. ☐
6. I understand that I will not refer to the name of my teaching program or working place during the data collection procedures. ☐
7. I understand the content of the research might be shared with other researchers for use in future research projects. ☐

8. I understand that the content of the research report written based on analysis of data derived from the interviews, questionnaires, and reports could be published in an academic publication, project website, media publication, teaching or training materials for use in University activities or public engagement activities

☐

9. I agree that my contact details can be kept securely and used by the researcher to contact me about future research projects

☐

10. I agree to take part in the above project.

☐

_____	_____	_____
Name of Participant	Date	Signature
_____	_____	_____
Name of Researcher	Date	Signature
Taking Consent		

When completed: 1 copy for the participant; 1 copy for researcher/project file

## **Appendix 10: Mirjana's Transcriptions**

### **Appendix 10.1: Transcription of Mirjana's Open Ended-Questionnaire**

#### **Part One: Work Experience**

Researcher: Could you briefly describe the assessment role you have been involved in.

Mirjana: Since I started teaching I have been involved in preparing all the exams, like not all the exams, but most of the exams myself, especially in the first place I worked for, I was the class teacher alone, so we were preparing together with the other team members all the exams that we are going to test students with and that I was involved in. I was not in the exam unit in my other workplace, but I was involved in assessing the students' speaking skills. I also worked as a speaking grader and also a writing grader in the writing committees. Then I was a team leader, and then I prepared all the writing assessments for the preparatory program and then, later on, I was assigned the head of the department in the first place, and I reviewed all the exams and all kind of assessments such as quizzes, projects, and others.

Researcher: So, you were the decision-maker when it comes to any assessment.

Mirjana: Yes. Planning, deciding on the deliverables like how many deliverables we will have for the project for example, how many parts we are going to have in exams, in the midterm, in the final. With the team module team leaders, I was involved in a variety of assessments.

Researcher: For how long have you been involved in this role?

Mirjana: Since I started teaching in 2003. So, it has been a long time.

Researcher: How were you chosen for this assessment role?

Mirjana: The thing is, in different workplaces, I had different roles. So first of all, for the first place I worked for, it was the school and teachers who prepared the assessment; we did not have an assessment unit. So, teachers used to prepare the assessments, all of them. So, you know, we had to prepare them all with the other members of the group. Furthermore, for the second working place, I was not involved in the preparation part because I was not in the exam unit. So, I was only involved in the grading part for the speaking and writing exams as a grader there, and of course, we were giving feedback about the questions, like we were receiving some feedback from the exam unit, so we were in cooperation with them. So, it was a separate unit, but they were involved in the

process. We were giving them our queries, and then they were coming up with the revised answer key. So, it was like a communicative approach to grading. And then in my final role, I worked in my final workplaces let us say. I worked as a team leader first as the writing model team leader and then “how was I chosen there”. Like it comes with the role itself, it comes with the title. So, in our context, the team leaders prepare the assessments with the help of the HOD, so, I was given this role as a team leader, and I prepared the quizzes, the exams, written assessments, assignments, tests. I was responsible for the writing module, and my final role was the head of the Department of the Preparatory Program. So, I was more actively involved in this role because I had to review all the exams, all the tests, all the assessments, all the assignments, whatever students are being tested on. I was involved in the decision-making process and also the grieving process and also creating the exam process. Everything I was asked to see, I had the chance to be involved in assessments based on the role I mean, and that role requires this involvement. The choice is not based on specific criteria related to assessment. It was role-based.

Researcher: Describe your assessment tasks in points.

Mirjana: I created exams, I reviewed exams. I also reviewed all the assessments, assignments, deliverables, projects, everything, presentations. I graded speaking exams. I graded all the writing exams, as well. I also graded quizzes and tests and everything that we were trained for grading as well. That is why, like I am mentioning those two like speaking and writing, we were trained to grade all these subjective assessments in my previous working context. Yes, like we had those standardized meetings, and then they gave us sufficient training about how to be as objective as possible when it comes to grading subjective assessments.

Researcher: Were you involved in creating rubrics?

Mirjana: I also got involved in creating rubrics for writing and also for speaking as well. Moreover, I was involved in the reviewing process. I also was involved in preparing the Presentations. I was also involved in communicating results for students. Like when I was teaching actively, I was communicating the results of the students myself after grading the exams. And then in my current role, I am just supervising the process like not getting involved because I do not have a class now. I was involved in giving feedback to students.

So, to summarise, I was involved in designing, decision making, creating and using rubrics, grading, communicating results to students and giving feedback to students. So I was involved in the whole assessment process on the different skills and different level in the four skills: reading, writing, speaking, reading, grammar, vocabulary, literally and everything.

## Part Two: Educational Preparation

Researcher: Please name your education qualification.

Mirjana: I have a master's degree. It is about teaching Turkish as a second language is a foreign language. I do not have certification specifically in assessment. I did a certificate in teaching, as well. It included assessments somehow. It was as part of one certificate that I earned for teaching in general. And. I took a pedagogical course for teaching skills as well. There was nothing devoted to assessment.

Researcher: During your pre-service preparation, before you started teaching, were you exposed to any language assessments preparation, something

Mirjana: Before I start working. No. Because my bachelor is in English language, literature. So, there was no module for assessment.

Researcher: During your in-service preparation. After you started teaching, have you been exposed to any assessment training before performing your assessment tasks?

Mirjana: We had this internal training. It involved all skills. I also did "TDC" (Teacher Developmental Course); they all had some assessments' preparation training but not specifically about assessment.

Researcher: So can you say that you learned by hand on assessments?

Mirjana: Yes. Yes.

Researcher: After you started your assessment task, have you been involved in any professional development related to assessment on the job. I mean, like workshops, training about assessment.

Mirjana: In my previous working place, yes somehow, not on assessment but on grading. We took nothing related to assessment design because the exam unit was involved in it. So, we did not need to have any training because I was not in the exam unit. However, for grading, we used to have like standardization meetings for the speaking and the writing parts of the exams to make everybody is on the same page. To make sure they are using the rubrics correctly and they are applying all the items there in the rubrics in the right way. So, these we called standardization meeting. It aimed to have some standardization among teachers.

Researcher: So, before you started grading, you were trained. Furthermore, for the other assessment components, were there any workshops, training?

Mirjana: No, not really. I wish we could, but it was just personal sharing personal experience, you know, between colleagues, I work hard to prepare this, how to prepare that? What would you do? My husband is also in this field, so I was discussing all sorts of things with him. He worked in the exam unit, so he knows certain things, knows individual specifications that should be prepared.

Researcher: So, you tried to work on yourself using external sources. Did you try to use the internet, online resources or support?

Mirjana: Yeah. I mean, like from time to time we searched for certain things. Yeah, for example, When I had this head of the department role, I wanted to know when we are creating rubrics and stuff like that, we went online and searched for others, you know, other universities, how they do it and other things. Of course, we did some research, but we were not provided by anything and training. I did it myself because you want to perform well. We know that the tests should be reliable and valid and everything. So, we were trying our best to make them valid and reliable. So, I was just again, I was consulting my husband, plus I was searching online. And do you know, we were discussing certain things with the team members in order to make certain things more reliable.

Researcher: So you depended on experience, on your instinct, your feeling about what should and what should not. Did you come to a book or a reference?

Mirjana: No, no. I did it intuitively based on my experience as a teacher and then as a student as well. Moreover, plus, like we were in those pieces of training and those qualifications, we were told about these things, and we knew how exams should look like what tests are, what they should test or how they should test. So, I had some background knowledge about it as well. Nevertheless, then again, from my colleagues, from the training, from the workshops we had, we used to have many workshops. So, every single person was talking about different things. So, you hear lots of things at the same time, and you know that it should be this way. Not only my own opinion. I would also base it on some knowledge that I gain somehow, not necessarily from a book or from training. It is from here and there let us say. For example, if you want to prepare a grammar exam and you wanted it to be more reliable. I knew that it is better to ask certain things in context, it is my opinion, but I knew it from the literature that it should be. It should be more communicative, and it is better if the students are exposed to this kind of questions instead of tests on multiple choice. Alternatively, for example, when you prepare a multiple-choice question, when you have the distractors, they should not be any completely wrong. They should be grammatically correct. So, I knew these things somehow. I do not know-how. I do not remember exactly which resource gave me all this information. It is a vast experience. Yes. You have been dealing with colleagues, with superiors. Yeah. If it comes by experience.



Researcher: Were you given parameters for the work you were doing, or you did it from scratch out of your mind?

Mirjana: We put the parameters. There were not any guidelines. They just asked us to do it. We did not follow anything. We came up with our own. OK. We followed a sample. Yeah, but. We were not given any guidelines about how to create an assessment. Yes, this is what I am saying. We followed a sample prepared by other departments in the institution. However, we were not given any training or guidelines or reasons why we have to prepare them in this way. So, we did it ourselves, we changed certain things, we applied certain things we thought you know, this was a challenging thought for our students.

Researcher: Yeah. You were learning from your experience; you did it by hand on assessment. For how long have you been doing this?

Mirjana: I mean, active for seven years as a team leader

Researcher: Did you feel that you need a sort of support, guidance, training, maybe a little bit of learning. By the way, when I talk about learning, I do not mean only theoretical stuff because not every theoretical thing can work, and others do not work. It depends on the context.

Mirjana: Yes, Of course. I think whoever is given this task should be trained about it. Like I always felt that. As I said, because I was experienced, I knew lots of things like my experience, by chance, by workshops, by training I once I went through before, through colleagues, you hear things, you search for certain things. I attended workshops. I attended certification programs, let us say. But. It should be like that. I knew certain things, I was basing it on my knowledge, but team leaders, for example, some team leaders. I mean, do they have the knowledge or experience. Do you know what I mean? Like we always had difficulty when the team leaders were changed because the new team leader does not have a clue about how to prepare an exam. So, you know, without any training, with our basic knowledge, based on our experience in this context, then based on what we have done so far, we were training whoever comes. We were trying to give him/her some training. So, it is too much on the teacher creating the assessments. This person is held responsible for the exams, assessments, for everything without receiving any training. We were in the same shoes before.

## **Appendix 10.2: Transcription of Mirjana's First Structured Interview (Checklist)**

### **Checklist**

#### **Language Assessment Knowledge Base**

##### ***Checklist Aim***

This open-ended checklist is meant to explore the participants' confidence with assessment knowledge base proposed by the (TALiP) framework not to evaluate their assessment literacy level. It is also meant to elicit information about the participants' views regarding the proposed base knowledge to understand and interpret their conceptions and beliefs about the proposed knowledge base and their need for this knowledge in their assessments contexts.

##### ***Checklist Layout***

The checklist is divided into seven main sections. Each section addresses one assessment knowledge component as per the TALiP framework and involves sub-items that represents the sub-skills of each knowledge as per explored literature.

##### ***General Instructions***

Participants are kindly requested to perform the following for each item:

1. Read the items under each section.
2. Choose only one answer for each item and provide at least one verbal justification for their choice:  
High Confidence (HC) Medium Confidence (MC) Low Confidence (LC).
3. For each item, the participants will choose, needed (N) or Not Needed (NN); however, they need to provide at least one reason for each choice verbally.

4. The participants can ask for clarification on the provided items if they are not familiar with them or are not able to interpret the intended meaning.
5. All provided verbal responses will be audio recorded.
6. The interviewer can explain details about any item if requested by the participants.

## 1. Disciplinary Knowledge and Pedagogical Content Knowledge

*Knowledge about language content and how this content is learnt and taught. Any Assessment starts with specifying what to assess*

Theoretical Knowledge	HC	MC	LC	Comment	N	NN	Reason for N/NN
1. Knowledge of second learning theories:  <i>How learners learn a second language</i>  <i>(e.g. Behaviorism, Cognitivism, Innatism, Constructivism, Information Processing, Interactionism, Multilingual Theories)</i>		√		My BA is not in education or teaching. I was exposed to these kinds of theories after graduation during my pedagogical training. My knowledge is not based on my reading or exposure to this knowledge, but it is based on my knowledge as a language learner. I built this knowledge based on my own experience as a teacher and how students learn because I have the chance to teach different age groups	√		At the end of the day, we will test what students learnt and how we can improve them. Assessments and curriculum are interrelated to each other. We cannot think of them separately. We have a curriculum working on how students learn and what they learn; we have to test this according. We teach students, they learn, and we assess this. It is like a circle. We have to reach the same point of view. Of course,

				and I could see what challenges they face and what works with them.			teaching methodology is essential, but first, we should focus on the learner.
<p>2. Knowledge of second language teaching theories</p> <p><i>How teachers teach a second language</i></p> <p>(e.g. Audio Lingual Method, Community Language Teaching, Communicative Approach, Natural Approach, Content-based Instruction, Task-Based Language Teaching)</p>	√			<p>I read more about these things because, in the educational system, these things are given more importance than learning theories or approaches. I worked on it during my pedagogical learning. I was taught how to teach, teaching styles, methodologies, and what we should adopt. I applied this in my teaching career.</p>	√		<p>We need to know how to approach learners. Once we know their needs, we will know how to deliver in the best way possible. We need to know where to focus on, how to focus, and what kind of approach we should adopt to reach our students more efficiently and help them learn. As an assessor, it is crucial as well because again, they are interrelated. If I want a teacher to adopt a communicative approach, then my test should be a communicative-based as well. If I ask them to speak and focus on fluency, not on the accuracy, I could not test accuracy in an exam. An assessor needs to work on the practicality or mechanism of how to link methods to assessments. This mechanism is missing; I need it. It should again be evident to all stakeholders, teachers, assessors, students. I believe this mechanism is a significant component of assessor knowledge.</p>

<p>3. Knowledge of learning strategies</p> <p><i>Kind of strategies that could be used with learners to assist in learning a second language</i></p> <ul style="list-style-type: none"> <li>▪ <i>Memory Strategies (Memorizing, Retrieving)</i></li> <li>▪ <i>Cognitive Strategies (Understanding, Verification, Deduction, Induction, Reasoning, Creation)</i></li> <li>▪ <i>Compensation Strategies (Translation, Avoidance, Gestures, Guessing)</i></li> <li>▪ <i>Metacognitive Strategies (Arranging, Regulating, self-directing, Evaluating)</i></li> <li>▪ <i>Social Strategies (Asking, starting conversations, cooperating)</i></li> <li>▪ <i>Affective Strategies (Lowering Anxiety)</i></li> </ul>	√			<p>We were taught about learning strategies and how to use them to make learners learn a language plus from my experience as a teacher also.</p>	√		<p>I need it as an assessor to address individual differences. We could test material in different ways. We should not base it only on paperwork. We can use projects. Ask them to create a video, PowerPoint or the same material can be tested differently as a reading test or as a writing test at the same time. We need to make our assessments more balanced. We cannot have one standard exam for all. It is for ethicality and fairness.</p>
--	---	--	--	--	---	--	---

<p>4. Knowledge of learning styles:</p> <p><i>Individual learning styles, which learners can use to assist in learning a language</i></p> <ul style="list-style-type: none"> <li>Physical Styles (Visual, Auditory, Verbal, Kinesthetic, Tactile)</li> <li>Communication Styles (extroverted, introverted, Solitary, Social)</li> <li>Thinking styles (global, analytic, reflective, intuitive, open)</li> </ul>	√			<p>I was taught, and I got them by experience. I had some workshops about learning styles in my working context. Also, I depended on my knowledge about how to learn.</p>	√		<p>Students have different styles, and we cannot choose one. We have to design our assessments based on individual needs and preferences. Assessments need to address different learning styles. We are dealing with human beings. We have to give chances to each learning. If they cannot succeed in one component, he/she can succeed in another. We need to give them more individualized changes in assessments. Assessors need to think about them during assessment preparation to come up with fair assessments that address different styles.</p>
<p>5. Knowledge of theories of language use (Language Models)</p> <p><i>Language theories specify what constitutes a language or the kind of constructs teachers teach and assess according to different language models</i></p> <ul style="list-style-type: none"> <li>Canale &amp; Swain Language</li> </ul>		√		<p>I do not know the names of the theories, but I know their components.</p> <p>I was not exposed to this knowledge, but I know by experience on which component I should focus when we are teaching or testing a language. I know them, for example, I know how important it is to know about the world</p>	√		<p>This is the content of the assessment we need to know.</p> <p>We need to know what we are testing, are we testing lexis, grammar or what.</p>

<ul style="list-style-type: none"> <li>✓ <i>Grammatical competence (knowledge of grammar, lexis, morphology, syntax, semantics, and phonology)</i></li> <li>✓ <i>Sociolinguistic knowledge (the rules of discourse), and (the ability to overcome communicative difficulties).</i></li> <li>✓ <i>strategic competence</i></li> <li>▪ <i>Bachman's Model</i> <ul style="list-style-type: none"> <li>✓ <i>language competence (grammatical competence, textual competence, Pragmatic Competence)</i></li> <li>✓ <i>knowledge of the world (cultural and personal knowledge)</i></li> <li>✓ <i>Strategic competence (assessment component, planning component, execution component that helps using the language for communication)</i></li> </ul> </li> </ul>				and how it contributes to the overall knowledge of the language, but I was never exposed to the theories.			
---	--	--	--	---	--	--	--

## 2. Knowledge of Assessment Purpose, Content, and Methods

*Knowledge of \*Why you assess, \*What to assess, \*How to assess.*

### A. Theoretical Knowledge (Declarative Knowledge Used in Evaluating or Selecting Assessments)

Knowledge	HC	MC	LC	Comment	N	NN	Reason for N/NN
1. Knowledge of language testing and assessments history  <i>Phases that language testing and assessment have gone through</i>			√	I do not know about it. I just know how I was tested and how things changed now.		√	We should focus on what we need and do now because language is a living thing. I do not need to know about how they did it in the past; we need to go forward. We need to focus on the situation now because methodologies, content, learning styles were not even analyzed at that time. It was a very teacher-based education.
2. Knowledge of language testing and assessment philosophies			√	I was not exposed to the literature about this knowledge		√	This is more important than the previous one. If I would critique assessments, I need to know the idea behind them; what is the main goal behind the assessment. To know also if the assessments are matching the claimed philosophy of the teaching context.



<i>Different philosophies behind assessment (for evaluation/for learning/for both)</i>							
3.Knowledge of the impact of adopting testing or assessment philosophy			√	Again, not exposed to it but I can guess them by experience.	√		I need to know what the effect of a specific assessment approach or philosophy to accept it or refute it. As a teacher, I need to know what I am expecting and the result of what I am using on my students.
4.Knowledge of major assessment qualities:  <i>(Authenticity, Practicality, Interactiveness, Fairness, Ethics, Impact, Washback)</i>		√		Based on my experience, but I was not taught. I just figured it based on my experience or our discussion as an assessor team.	√		If you are preparing an assessment or even if you are just exposed to one, we need to know if it will work. Knowing the qualities can help users can judge, critique, give feedback. I believe all teachers need to possess this knowledge even if they are not assessing/designing. They need to receive some training on assessment qualities.
5.Knowledge of assessment types  <i>(Traditional/Alternative, Norm-referenced/Criterion-referenced, Summative/Formative, Externally</i>		√		I know and do many of them, but I am not familiar with the literature or terminology. I am confident, based on my experience or varied	√		As a teacher of classroom-based or criterion based ones, but standardized ones, I do not believe I need to know about them if I'm not involved in creating similar ones, but we need to be aware of the standardized ones because some of our students are involved in such assessments, so we need to guide them if they need help. As an assessor, I have to check their types of

Mandated/Internally Mandated, Standardized/Classroom)				resources. I did take the test myself.			<p>questions and how they are testing out there like IELTS and TOFEL. I need to critique them. For example, they are based on a very short-time assessment, and I do not believe they reveal the skills in the right way. They create anxiety; they are not individualized at all, and they are trying to put everybody in the same box. I still believe that we have to be there, but in our context, it is different; our assessments are more individualized and varied.</p> <p>I need to know them to say I am not going to do this. We need to be armed with the tools to face them.</p>
6.Knowledge of assessment methods:  <i>(Tests, Portfolios, Performance Assessment, Self-assessment, Peer assessment, Checklists, Interviews, Questionnaires, Role-plays)</i>		√		Based on experience	√		It is definitely needed. These are our tools. I need more skills in designing them.
7.Knowledge of major steps in language test development:		√		I was not taught these things. I cannot name them, but I do them while designing. I connect the items to the	√		If I were trained about such things, I would have been more confident about them. I could have guided the assessor and teachers more. This knowledge would

(Test Purpose, Construct Definition, Content Specification, test Specification, Item writing)				learning objectives, but I think it is not enough. I am doing them intuitively. It should be more advanced. "We follow the model."			have been reflected in the production of assessment more professionally.
8. Knowledge of cognition taxonomies required for different test items or tasks <i>(Bloom's / Webb's)</i>		√		I know them by experience. I cannot say I did this because I followed this taxonomy. We are not that professional. I cannot say I am a tester or an assessor because I was not trained, taught, supported. I was just given a task, and I dealt with it myself. I feel it, I do it but not in a standard/formal/ manner. What I do I feel is right, but it could be much better if we had some kind of training or if we could relate things to one another.	√		They need to be applied to the teaching process first, so we know that the student reached a particular cognitive stage so we can assess based on them.

## B. Practical Knowledge (Skills, Procedural Knowledge Needed for Creating or Designing Assessments)

Skill	C	MC	NC	Reason	N	NN	Reason for N/NN
1. Identifying and stating language assessment purpose	√			By experience. Already set by the programme.	√		I have to know the purpose of my assessment.
2. Writing goals and objectives of instructions and consequently their assessments	√			We critique ourselves. We do it as a group in the form of discussion as an assessor team. We depend on our judgment. It is already set by the programme.	√		We cannot assess them based on something we have not taught. It is not fair.
3. Aligning curriculum objectives, instruction, and assessment	√			My own experience because I worked on the curriculum department, and I know about the European framework, and I know about the level of students, what they should do by the end of a critical stage. As a team, we follow a plan prepared by another institution and have samples. We were just given this, and we did it in our way.	√		It is mandatory before exam design.

4. Defining the language construct(s) an assessment will give information about	√			Through experience	√		Definitely needed.
5. Utilizing alternative means of assessments to make decisions based on substantive information	√			It is adopted in my teaching context, and I also believe in it. I initiated some more alternative ones based on my initiatives.	√		After all, it is a preparatory program for at last one year. It depends on classroom-based assessment, so we need alternative ones not just exams.
6. Constructing test specifications/blueprints to design parallel forms of a test		√		I use a simple one for reading as an example. It specifies the level, the duration, type of questions, number of questions, the percentages. I did not initiate it. It from teaching context, It came with the package. I changed certain things because it came from another department.	√		It is essential because it makes assessment more standard and formal.
7. Writing selected-response items such as multiple-choice-, true-false, and matching	√			By experience	√		Of course

8.Designing constructed–response items (for speaking and writing)	√			By experience.	√		Of course
9.Constructing well laid-out and perfectly legible items/tasks	√			By experience	√		Of course
10.Providing clear, explicit, unambiguous instructions	√			By experience	√		Of course
11.Providing examples to make candidates familiar with assessment techniques	√			By experience	√		Of course
12.Including many items to enhance scoring reliability	√			By experience. We use many items in the tests.	√		Of course
13.Writing test syllabuses to inform test users of test formats where applicable		√		We do not use anything formal. We prepare students in the class for test formats so students can get ready for the type of questions.	√		To prepare students
14.Designing assessments that are valid not only in terms of		√		Through experience	√		

course content but also in course tasks							
15.Designing assessments that are reliable, authentic, fair, ethical, practical and interactive		√		My experience	√		Of course
16.Incorporating technologies in assessing students	√			We use some platforms, Turnitin, for example, for students to submit their deliverables. I worked on it a lot. We were in a stage where students did their writing exams on the computer, but then the institution did want it, but I felt very proud to do it because students used to have two drafts and we used to give feedback, and it made life much easier for the teachers to grade those essays online and to give feedback online, and students could see easily see and address our feedback.	√		Of course

### 3. Knowledge of Assessment Grading

<b>A. Theoretical Knowledge (Declarative Knowledge)</b>	<b>HC</b>	<b>MC</b>	<b>LC</b>	<b>Comment</b>	<b>N</b>	<b>NN</b>	<b>Reason for N/NN</b>
1. Knowledge of grading rational ( <i>Norm-, Criterion-Referenced</i> )		√		Experience	√		Of course
2. Knowledge of methods that change scores into meaningful composites  ( <i>Points, Percent, Grades, Proficiency Levels</i> ).		√		Experience	√		Of course
3. Knowledge of necessary linear scaling, weighting scaling, weighting components, and precision of results.		√		Not in details. I do it by experience.	√		At least the basic knowledge
4. Knowledge of scoring techniques for objectively-marked testing		√		Experience	√		We have to be objective
5. Knowledge of the nature, purpose, and design of scoring rubrics for subjectively marked tests		√		I design them, but I believe they should be given more importance. We should have more standardization meetings.	√		They should be graded twice from two different perspectives. I need more structured training



6. Knowledge of consistency and moderation principles for judgment-based assessments.			√	Basic knowledge	√		I need more
<b>B. Practical Knowledge (Procedural Knowledge)</b>							
1. Constructing scoring schemes that quantify learners' performance into useful information for decisions about learners, classrooms, schools and districts		√		I do it	√		Of course
2. Designing scoring keys with explicit instructions for closed-ended responses (right/wrong, checklist, multipoint methods including rubrics and rating scales).	√			Yes we do that	√		Of course
3. Designing scoring rubrics with explicit instructions for open-ended responses (holistic, analytic, primary trait scoring)	√			We do the general ones.	√		Of course, but we do not do that. We need to be more detailed. We assume teachers are grading based on the rubrics, but it includes elements of subjectivity. I guess we need to have some training for the grades.

4.Calculating consistency of scores/ standard error of measurement to assess scoring reliability for both closed-ended and open-ended responses			√	We do some cross-evaluation.	√		
5. Calculating scorer reliability through scorer reliability co-efficient to quantify the level of agreement among scorers			√	We started doing that.	√		It is more than needed.
6.Designing training workshops for scorers on acceptable approved responses			√		√		

#### 4. Knowledge of Feedback

<b>A. Theoretical Knowledge (Declarative Knowledge)</b>	<b>HC</b>	<b>MC</b>	<b>LC</b>	<b>Comment</b>	<b>N</b>	<b>NN</b>	<b>Reason for N/NN</b>
1.Knowledge of purposes and principles of feedback	√				√		
2.Knowledge of different types of feedback (descriptive, evaluative, supportive) with their respective functions		√			√		
3.Knowledge of the strengths and weaknesses of various kinds of feedback (task, process, metacognitive, and self-oriented) in facilitating student learning		√			√		
<b>B. Practical Knowledge (Procedural Knowledge)</b>							
1.Setting educational goals and constructing tasks that will move learners towards these goals	√			Experience	√		

2.Communicating to learners the criteria for success through different communication modes like telling, showing, and having learners discover	√			Experience	√		
3.Showing learners the gap between their current stage of development and the next/final stage of development		√		Experience	√		
4.Intervening in the learning process either by taking an “interventionist” or “interactionist” role			√	Experience	√		
5. Scaffolding their learning through different mediating exercises “graduated prompt”, “testing the limits”, and “mediated learning experience”.			√	Experience	√		
6.Using internet resources such as online tutorials and adapt contents to address students’ particular needs		√		Experience	√		
7.Using assessment methods to monitor, learning by collecting formal data (tests)		√		Experience	√		

and informal data (observing) of students' language development							
8. Providing encouraging feedback that is meant to modify the learning process not just a score		√		We do not have this kind of practice.	√		
9. Monitoring, recording, and reporting student language development		√		I never used this		√	For now, not but maybe in a different context.
10. Evaluating how well learners have succeeded in achieving the final learning goals		√			√		
11. Improving instruction based on assessment results and feedback		√			√		

## 5. Knowledge of Peer & Self Assessments

<b>A. Theoretical Knowledge</b>	<b>HC</b>	<b>MC</b>	<b>NC</b>	<b>Comment</b>	<b>N</b>	<b>NN</b>	<b>Reason for N/NN</b>
1.Knowledge of benefits of involving students in assessment	√			Through teaching experience	√		Students are our material. They have to intervene in the assessment procedures. They should not be excluded. If we talk about peer assessment, they learn more from one another. It is better when they work in groups and peers. It is less stressful, and they work with less anxiety.
2.Knowledge of strategies of using self- and peer assessment in different assessment tasks		√			√		
<b>B. Practical Knowledge</b>							
1.Training learners to effectively participate in assessments		√		I can do the training. We do not involve them in creating or grading the assessment or grading each other peer assessing or assessing	√		It could be much better and fair.

				themselves. In this context, we do not practice it.			
2.Using self- and peer-assessments to promote learners' ability to notice their learning problems and gaps in their target learning goals	√			I do it in the class but not for formal assessments.	√		
3.Providing learners with criteria used in judging their work and make them use it to self or peer assess their work and their peers work		√		We do that	√		
4. Encouraging learners to produce their rating criteria in groups and use them to assess one another.			√	I can do it, but I have not tried it.	√		However, students need to be guided first and trained.
5.Training learners to use portfolios, diaries, continuous cards, digital audios, video diaries, online blogs where learners are expected to collect samples of work and commentary to monitor their learning process and keep records of how their work improves and develops			√	I have never tried it. I have never had a portfolio work.	√		It is definitely necessary.

## 6. Knowledge of Assessment Interpretation & Communication

A. Theoretical Knowledge	HC	MC	NC	Comment	N	NN	Reason for N/NN
1. Interpreting data related to test design such as item difficulty and item discrimination		√			√		
2. Interpreting the concept of reliability in language assessment:  <i>(Dependability, Item analysis, Reliability Threat)</i>		√		I know about item analysis. We did it internally no one asked for it, and we made some changes based on it. I am not familiar with reliability threats.	√		Of course, needed but we need some training on them.
3. Interpreting validity in language assessments  <i>(Construct, Content, Criterion, Consequential Validity, and Validity as Argument)</i>		√		I know about them but again based on experience.	√		Of course, needed
4. Interpreting data from large-scale tests:			√	We only collect the data, but we do very basic ones and	√		I do not need a lot. In my context, I need some basic ones, the average, and comparison between groups. Maybe



(namely means, modes, medians, bell curves and can calculate them)							someone else would do it. It does not match what we do in classroom assessments
5. Inferring students' strength and weakness based on collected data to communicate it to different stakeholders		√		We do it, and we work on it.	√		
<b>B. Practical knowledge</b>							
1. Calculating item difficulty and item discrimination for close-ended items to obtain reliable scores		√		I do them in a very simple manner.	√		I need more about them
2. Investigating facility and discrimination indices statistically			√	No		√	I do not need it someone else would do it
3. Calculating test/item reliability, calculating inter/intra-rater reliability		√		We do some item analysis, and we see if things cannot work.	√		I need to know more about it

4. Investigating validity using statistical procedures			√	Not very confident because I do not use that much and the institution do not require it that much		√	I need to more about it dealing with statistical information, mostly if I worked in a different context.
5. Articulating interpretation of norm- and criterion-referenced assessment results to a variety of audience: student, school, directors		√		Yes, but only criterion-referenced ones or classroom-based ones.	√		Classroom-based ones
6. Communicating interpretations using different reporting tools: Report cards, documents, criteria, guidance counselling.	√			We do them on Moodle as per context requirement.	√		We do not need official ones as per my context
7. Using software such as Statistical Package to communicate results			√	Basic ones		√	For classroom-based ones, we do not need it, maybe for more standardized assessments.
8. Participating in committee or school-wide discussions about reforms in assessment-related issues, curriculum, materials, grading policies, accountability	√				√		

policies, school evaluation and school evaluation.							
9. Having the skills to coach students to analyze their assessment results, track their learning, communicate about their learning, and plan next steps.		√			√		

## 7. Knowledge of Assessment Ethics

A. Theoretical Knowledge	HC	MC	LC	Comment	N	NN	Reason for N/NN
1.Knowledge of meaning and implications of critical language testing: power, ethics, and fairness			√			√	I do know, and I still prepare. I prefer to do it intuitively. I do not need to study this we have these ethics already
2.Knowledge of how to work towards equity, non-discrimination, inclusion, and social justice		√		I know things intuitively.		√	This is intuitive knowledge
3.Knowledge of codes and concepts of professional morality		√				√	I know enough
4.Knowledge legal and ethical responsibilities concerning the use, storage, and dissemination of assessment results			√			√	It is done based on your experience or your institution requirement. This is too theoretical. It is already set up, and you just follow the rules. It is not my role as a designer.
<b>B. Practical Knowledge</b>							

1.Observing guidelines for ethics concerning language assessments	√				√		
2.Treating all students or users of language assessment with respect	√				√		
3.Providing assessment practices that are fair and non-discriminatory through assuring distributive justice	√				√		
4.Avoiding construct-irrelevant controversial, inflammatory offensive or upsetting test material	√				√		
5.Adopting transparent language assessment practices by informing students of what, how, and why of assessments	√				√		

6. Involving other colleagues in assessment decision making and practices	√				√		
7. Implementing democratic practices by accepting and being open to constructive feedback from colleagues	√				√		
8. Using a range of assessment approaches and multiple measures that allows students to show their knowledge	√			We do that.	√		I do not think I need to read about it is by practice.
9. Defining a clear purpose for assessments, develop specifications, evaluate the content and conduct a field test examination		√			√		
10. Using tests' scores ethically through providing evidence of fairness via statistical procedures such as validity and reliability			√		√		

11.Clearly and honestly informing the inferences and decisions that derive from scores in assessments	√				√		
12.Using assessment results for feedback to influence language learning	√				√		
13.Evaluating the kind of washback that assessments can have on learning, teaching, curricula, and institutions		√			√		
14. Judging the consequences (intended or unintended) stemming from assessment in own context through examining "consequential validity."			√		√		We need to judge the impact of our assessments.
15.Implementing democratic language assessment practices by giving students the		√			√		

opportunity to share their voices about assessment							
16.Critiquing the impact and power, standardized tests can have			√	I am not aware of them.	√		

### How do you evaluate the list?

For me it a good reflection practice on what we are doing as assessors and teachers, what we should be doing, what we are missing. It is very detailed. It is feeding me somehow with information about me as an assessor about the institution. I feel we need more standardization, more guidelines. I feel I am missing out certain things. It raises my awareness. It is perfect. It is very well prepared. It is good to review what we are doing her internally. It is for y self. It is good to raise our self-awareness.

**I noticed that when I came to the practical ones, you mentioned that you were highly confident, but when it came to the theoretical terms, you were not.**

Yes, I would instead be trained about things which makes my argument stronger if I have to defend my practices. At the same time, I do not need that different theoretical knowledge because it comes by practice, and it depends on your context need to have. It needs to be flexible because we can never have set rules about preparing assessments, but at the same time, I think people should be prepared before taking such assessment tasks and role. I learned by hand on practice, and it took me years to say that I am highly confident about some of the things.

**You mentioned that you need more in many stuff. There is a gap. Something is missing.**

I would be more confident if I knew the theoretical knowledge self even. When I am saying that I am highly confident in practice, and I am doing all of these things. Still, I would appreciate the opportunity of knowing the knowledge I am basing these things on. I am only doing these things, but it is in the air. I cannot say I do things because this is how it should be. I am doing these things because I feel they are right intuitively. I know as a teacher as



a learner as a test taker myself, it should be this way but was not taught. I would instead be taught myself. I would have been able to connect theoretical knowledge and practice. I would have closed the gap.

What was missing in your role?

We were just given the task and specifications. We had samples, but we did not how to prepare an exam, what to focus on or how we should test things in different ways. I definitely would have appreciated the opportunity of having a practical knowledge on where to star, how to start, why we are doing these things, how we can improve it. I am not talking about the history of assessment, how people did it before, or language should be assessed this way, but real guidelines, a checklist. I want to go to the thinks that I ticked as highly confident in your checklist, and I want to ask a person to give me a training about why I am doing these things or if a doing the right things. I do these things. This is the right things I know, or I feel it is right. I need to relate it to the literature as well. I think this is missing here. We were thrown to experience only instead of someone helping or supporting us for the process. We were just given the task to deal with it, and we dealt with it from a teacher point of view from a learner point of you, but we luckily we had enough training as teachers. I was a well-trained teacher. I know what was going on classes so I could relate to my assessments, but not everyone has this much experience or training in teaching, so why not. I told you I would definitely need training in teaching or assessing/preparing/designing.

**If you were given a choice to be given and assessment course or training in your context based on the assessment requirement in your context.**

Training of course, why would I need theoretical knowledge or how people are assessed in other parts of the world or different types of assessments that I am not using. When it comes to anything, workshops, training, conference, everything should be based on practical implementation, how to do it, show me, how to do it. Otherwise, I will listen and say ok interesting. They are applying the most assessment practices in Japan; let us say so what. I would say wow, but it is not useful to me in my context. I would instead focus in what I am doing here and ho I can make it more ethical, reliable, better, valid, what I need is training based on my context, not a general assessment course.

**What about knowledge in more standardized tests and their statistical procedures? Even if you are not involved, do think you need this knowledge? Maybe you are not using them now you might need them late on.**

Basic knowledge would be enough.

As language teachers, we have to be familiar with what is going on in this world because these standardized tests although they are not for me very fair, to be honest, they are standard and used internationally we have to be familiar with them. We have to know their tools because, in our context, we are accepting their scores as a substitute for taking our programs. We are competing with them in a way. Of course, we are not preparing them for those tests, but we have to understand the meaning of these scores or these levels, so I know my goal and prepare students accordingly, where am I trying to carry the students, to which level. At the end these exams are prepared by professional, they are checked, there are all these types of formality, reliability, validity, fairness, although I said they are not individualized and too rigid, we need to know what is going so we can critique, compare and may be applied or refuse some of their tools and practices.

### **Appendix 10.3: Transcription of Mirjana's Semi-Structured Interview**

**Researcher:** This is the third stage of the research, which is the semi-structured interview here. I am going to have around 14 questions. You can talk as much as you want about each of them. The only thing is that for every aspect, we need to know what happens in your context. Moreover, after you mention what happens, you need to mention your belief of how things should happen in relation to this aspect. So, first of all, why do you assess students in your teaching context? For what reason?

**Mirjana:** In our context, we have the students, almost all of them, for one year in the program, and we are preparing them for their academic studies and their undergraduate studies. Furthermore, in this place, the language of instruction is English, so we are trying to carry the students from one level to another level, which is enough for understanding and studying in the English language so they can continue their studies without any problems regarding their language. That is why throughout this journey, we have to assess them somehow to see if they can do the necessities, the requirements of that particular level. So, we have to try to reach B2 level roughly at the end. Of course, there are some specifications about that level based on the common European framework, so the main objective of this program is to, as much as we can, help somehow students reach that level, so they do not have any difficulties dealing with their academics in English. Moreover, in order to assess the performance, we need some different types of assessments so that we can see the results. I mean, the students either pass or fail this program if they get a percentage of the overall score, 70 per cent of the overall score. So basically. I mean, the reason behind assessments is teaching and then assessing what we are teaching. Plus, we need to see if they have reached that level.

**Researcher:** So it is assessment for achievement and evaluation. What about your reflection about this?

**Mirjana:** I mean, I think in every program there should be assessments. We have to have assessments. We should not have only one assessment but different types of assessments. However, like I think the problem in this context is not the assessments, but the level of assessments because we have standard levels like we have one level for the first semester and one level for the second semester. If they cannot pass the first, they can repeat the first level in the second semester, but two levels are not enough. So, that is why, like if we had more levels based on students' language skills, then we would adapt the assessments accordingly as well as the curriculum accordingly. I think the problem starts with the placement test. In my opinion, all the foundation programs should have a valid, reliable placement exam through which students are placed in the correct level and the curricular method. All assessments are prepared according to their level.

**Researcher:** So maybe everything should be based on need analysis first and then when it comes to the purpose. As far as I got it, it is just for assessing the achievement of specific goals or objectives.

**Mirjana:** Of course. Because we have different types of assessments, in the end, the outcome is either pass or fail. So basically, they are collecting these points from different types of assessments maybe we have in total ten to fifteen assessments in one

semester. So basically, the percentages are very low. And then they are collecting those points to reach the 70 per cent overall score. However, some assessments are used for checking learning as well. They are acquiring some skills like presentation skills and also writing skills. And we are taking it to step by step. We are not asking them to write essays straight away, so we are assessing them and then see the problems and then teach more and then check the learning, and then we expose them to different types of assessments, so it is not only one assessment that they fail or pass.

Researcher; Are they summative or formative assessments? Do you have this kind of formative assessment, like just assessing to know their learning level, to know the weakness, strengths?

Mirjana: we don't have formative assessments, but we have pop quizzes in the classes. Teachers are preparing them just to see if they have learned or not, so t they can prepare the next lesson accordingly. However, they are not counted for anything, even physically.

Researcher: So, you are OK with let us say with the philosophy of assessment in your teaching context.

Mirjana: if you ask me. I think I believe in more assessments, more frequent assessments, then having big exams with high percentages so that we can see the process and we can act upon it. And if we have like short quizzes that can make us see where they are, and we can take action accordingly instead of waiting until the midterm, which is in the mid-semester and it could be too late for some students to catch up after that period. We do have quizzes. Nevertheless, I would rather have them more frequently.

Researcher; what is the source of assessment in your teaching context? Are they internally done or externally mandated tests?

Mirjana: We are doing them internally.

Researcher: Are you with or against this?

Mirjana, I am with internal mandated assessment because we have to have assessments that suit our context and our learners' need. We cannot rely on external resources to assess our students with their different levels, backgrounds, language skills, and unique specific culture. So I think the more individualized assessment sounds, the better they are because it is a sensitive context.

**Researcher:** what assessment methods are used to assess learners' skills in your teaching context, are you using tests, interviews, checklist, projects presentation? Can you talk about the kind of assessments, assessments you get?

**Mirjana;** we have tests, which have like multiple choice, gap filled questions and things like that, so. We also have quizzes. We have midterm, and finals, which are similar in style and question types. So the students have to sit for the exams and answer the questions. We have the writing part in the mid-term and final exam, where students have to write a piece of a paragraph or a text or an essay; it is sometimes a part of an essay, not a full essay. We have project work, which involves both writing and speaking. So, the first three deliverables are usually writing, and the fourth and fifth deliverables are speaking, but we are stuck with the presentations, which are usually a group presentation or an individual presentation. We also have a separate presentation from this project work. Like we can say that every semester we have two types of presentations, graded presentations. Students are trained of course before that so that they can give the presentations. What else we have in-class assignments? This is a different thing. Maybe not many programs are applying this. In class-assessments are done in the class. Not in an exam format, students are basically in their comfort zone, in their classrooms with their friends. These are graded, and it is in a very relaxed atmosphere, so students do not feel anxious about the exam. They do not have very high percentages, but it is good to make them and produce something based on the language they acquired so far. Moreover, I do it in a friendly atmosphere. So basically. These in-class assignments are less formal, but I think they serve very well in terms of seeing the weaknesses of the students and work on them before the exams.

**Researcher:** As far as I noticed, you have a variety of assessments. From your point of view, Are you happy with this?

**Mirjana:** Most of the assignments I am happy with, but the only missing thing for me is the speaking exam. I think there must be a speaking exam if you are testing the language. We have presentations, graded presentations, as I said, but it is not the same as a speaking exam. Speaking exam should be between the students. They are given a topic and they talk about this topic, and then they are graded based on their fluency mostly and their accuracy as well.

**Mirjana;** you have a problem with the presentation because you feel that it is not natural and maybe not authentic.

**Mirjana,** it is prepared, and the students are prepared, and they are presenting, this as a different skill than speaking skills. Because we do not have this participation grade, we cannot evaluate students for their speaking skills. So I think there should be a very organized structure for speaking exams where students come in pairs or groups of three and then they are given a topic in a group to make it less stressful for the students.

Researcher As far as I understood, you are happy with the variety of assessments. However, maybe in some of them, you want to manipulate them in order to make them may be authentic to reflect the real competence of students

And who is involved in assessment design? Is it a specifically chosen committee or all language instructors involved in the assessment process?

Mirjana: only a committee. Let us say involved in the assessment preparation design. They are the team leaders and the head of the department. These people prepare the exams, and the head of the department reviews the exams, and they are stored and printed without having the instructors look at it.

Researcher; are you happy with this.

Mirjana. The other instructors are not entirely excluded from the process because they know the format of things, and they know the types of questions; they are given the instructions to mark the exam. I think it would be a better idea to involve them more in the preparation and designing processes. Because they are the ones in class, they teach the course themselves, and they know their students very well; they know what they can achieve, what they cannot achieve, what they focus on or do not focus in the class. Team leaders also teach, but they are only exposed to one group of students, but if the teachers are involved, then there will be some discussions, more people will be involved. It will be more individualized and more customized based on the needs of students. A better idea about the majority of students and the burden and the responsibility is going to be shared. Now the load and also the responsibility are on the team leader; it is too much.

Researcher: Do other teachers want to be involved?

Mirjana: they want to get involved for sure. Most of them, maybe not all of them, but most of them would like to get involved in the process, but the decision is not ours. But we involve them somehow in the quiz preparation, the informal ones, the informal ones they are not involved.

Researcher: Can I ask, even though you have a variety of assessment, do you still feel that the formal tests are still dominating?

Mirjana: Yes, because the percentages are right on that. That is why I think maybe we should have, as I said, more frequent quizzes or in-class assignments or project works or writing task. Then maybe a mid-term or a final exam.

Researcher: Do you feel that this context is more into collecting scores?

Mirjana: We have attendance grade, project grade, and presentation grade and in-class assignments. We have different types of assessments. I would love to have more of this kind of assessments than having still one exam, especially the midterm exam. OK, maybe we should have something at the end of the semester to see the level of students and if they can manage the next level. Because we do not have a placement test, the main problem starts there, I believe. That is why we cannot satisfy the needs of the students in terms of assessment.

Researcher: maybe you are more into formative ones throughout the semester and having one summative at the end.

Mirjana Maybe this will help us to see how they learn, what they learn, and if they need to learn more about something, so we can take actions during the process until we reach that summative.

Researcher: You have mentioned something regarding the attendance and the percentage given to it. Are you with this?

Mirjana No, it is not an assessment. It is just attending a class. I would not penalize or award the student for attending a class.

Researcher: What other grades you give students, and you feel that this is should not be part of an assessment or counted towards a final grade of assessing a student's performance?

Mirjana: Sometimes, the deliverables of some projects. Like, for example, we are asking them to bring the visuals, and then we grade the visuals. So maybe there is no need for that. So, we can reduce the number of deliverables instead of giving so many deliverables and then grading those deliverables. Maybe there is no need for that one because the percentage is high while the only thing the students are required is bringing the visuals and then they would be graded for that. So, it is not an assessment, but it is part of the big project. I believe it should be put into pieces as much as possible, but at the same time, I fear that students will feel overwhelmed. Our students in this context are not aware that this will help them to pass in the long run. They are just anxious about the exams, so they see assessments as threats. So, they might disagree with this idea of having more small types of assessments. I feel it is based on their background. Maybe they have come from such a system where they have to take an exam and then pass or fail. So, when you expose them to different types of assessments, they see them as a threat, they do not consider them as like part of their learning. I need to do well perform. I need to see myself. I need to test myself. They do not have this skill yet. They are not able to self-evaluate.

Researcher: If we mentioned that there are some grades or scores that are not revealing their competence actually or added for the satisfaction of students like, for example, 10 per cent attendance. Again, it is affecting the validity of that assessment.

Mirjana: I would consider assessment out of 90 per cent. Because I cannot count attendance, it does not. We do not have a say in that. They are either given the grades or penalized because they have not attended the class. Maybe they will do better in an assessment. Even for the students who do not attend the class, it is not fair to penalize them for not attending as well. It is not fair to give a grade for just attending and doing nothing. So basically. I think it should not be there. This is not a valid way of assessing.

Researcher: And when it comes to exams, the constituents of exams like what kind of things do you put in an exam?

Mirjana: We have like multiple-choice questions. We have reading texts. We have four parts. We have vocabulary, reading, grammar and writing. So, in the reading part, we have two texts, one of them is true/false, the other one is multiple choice.

Researcher: Do you apply peer or self-assessment principles and approach using your teaching context? Like, for example, do you have forms or schemes to give the students to have the chance to assess one another and consider it or assess one oneself and consider it?

Mirjana: No. We do not have it formally. I mean, of course, those features are there especially when teachers are teaching the presentation skills, they give the rubrics to the students to grade each other, to give feedback to each other. We do these things in class, but it is just for practice not for real grades.

Researcher: Do you think teachers and students take these things seriously?

Mirjana: some of them do, but some of them are very harsh on their own, their peers.

Researcher: What about self-assessment?

Mirjana: Self-assessment. I do not think it will work here. It is based on their background. I do not think they are used to assessing themselves. Once it comes to assessing themselves, they will all give full points to themselves. They will not be critical. They do not reflect on themselves; they do not have the skill. They are not taught. Maybe if they are taught, if they are trained on how they could evaluate themselves, yes they would understand.

Researcher: Do you believe that you might once try.



Mirjana: It could be applicable. However, as I said, because their main concern in this context is the grades, even if they know that they do not deserve the grade, they still want to take the grade. That is why I think it would be a difficult task to break this habit of. Students are coming from such a system, that the grade is the most important thing for them and it does not matter if they performed well or not as long as they get a good grade. So, it is difficult to train these students in this context about evaluating themselves critically in a very fair way because, in the end, they know that they would not be graded for their self-assessment and it will not valid; they will not be harsh on themselves. They can be very generous with that stuff. So it is difficult to treat these students, but of course, they can be tried. It an important skill. In my opinion, it should be part of the assessment. For example, in the writing assessments, we have checklists for them to check after they finish writing. They are not grading themselves, but they are checking if they have completed tasks. So, it is kind of a self-assessment as well. But again, as I said, students should be trained, they should know how to look at certain things in their performance.

Researcher: Maybe if students are prepared. If somebody talks to them about the benefit, these things should be introduced to them; they should be trained. Teachers/assessors should have the skills and the tools and the techniques of how to do it and maybe pass it to them.

Mirjana: This is the thing. If you go back to the types of assessments, I think for writing, we need a portfolio of assessments because in writing I might teach students how to reflect on their writing, it might give them a chance to do self-evaluation about their performance and that because they will have a chance to correct. Because writing is a process, you cannot judge one paragraph. And we cannot give the grades to that paragraph and whatever we say after this paragraph, whatever feedback we give to students after this paragraph will not work as much as it is a process, like if the student knew that he would improve his grades taking care of this feedback and then making some changes based on the feedback, then this will work. Yeah. And then they will start to be critical about themselves as well. So, they will see their mistakes. Because there is hope that it might get better and it might improve. So, I think writing should be a process more than one final product.

Researcher: And now we have come to the most crucial part. How are the language skills assessed in your teaching context? Let us start with reading; how is it done?

Mirjana: We test reading in midterm and final exams and also in quizzes. We have in the midterm and the final exams two reading texts in each, and one of them with true-false questions and one of them is with multiple-choice questions, and we have a bonus question which was required. This is how it is tested. And we also have quizzes, reading and vocabulary quizzes let us say, the students are exposed to these texts before the midterm exam. It is a kind of preparation for students to get ready for the big exam formats, that same type of question because we follow the same type of questions in the quizzes as well.

Researcher: Are you happy with this kind of assessment?

Mirjana: I think it has to be some kind of reading where students have to use their skills. The question types can change. I do not say like the questions have to be multiple choice, or they have to be true or false. It could change. I mean, we need to focus on their competence in comprehending the text like assessing skimming, scanning skills. They sometimes get the answers by chance.

Researcher: because they can get the answers by chance. What do you suggest?

Mirjana: We can have open-ended questions. But again, it will give different opinions then it will be difficult to grade and then will need some kind of standardization for grading as well so because there will be so many queries about the answer keys. But there are other ways to check the comprehension of the text, so it does not have to be multiple choice, it can be like, as I said, open-ended questions. It could be like matching questions. They can match the main ideas with the paragraphs. Unfortunately, the current structure is rigid and enforced on us. If I would choose, I will change it. I would not have a standard type of reading. I will not have only matching questions. I will change the questions in every exam based on the text itself. So as long as we are testing the same thing, it does not matter what type of question we are testing it with. This is always my argument because in the book for example we have insertion questions, we do not have them in exams. We are teaching referencing, reference words and everything, but we don't test these skills. Now, after we have realised that they need more practice, we can say that we are preparing them for exams. But as I said, I think teachers plus the students are focusing mostly on those types of questions for the sake of scores.

Researcher: But you are teaching many reading skills, but you are not testing them.

Mirjana: yes, the book has different types, like making inferences, Let me put it that way, If we have a reading text we should have questions testing comprehension, others testing scanning, skimming, inferences, referencing. So basically, we should try to address the skills that we taught in class. But for me, it does not have to be multiple choice only. It also restricts the exam designer.

Researcher: We have been talking about this Bloom's taxonomy and all these kinds of things and the different levels of thinking. Do you think that your current reading assessments address critical, evaluation, reanalyzing, re-ordering and many other learning skills?

Mirjana: No, of course, not to that extent. To be honest with you, I do not think that it is also taught in the class. But the problem is level. When we tell the teachers to focus more on other reading questions and to start a discussion about the reading. They always come up with the same excuse. There is no time, and we always have to follow the syllabus because we are taking students from an A2 level to B2 level which is a long journey, but we have minimal time to complete the story. So, again I will come to my starting point. We should have a placement exam. So if we had the placement exam, then everything would be different, the levels would be different. Everything would be smooth. Then we can teach and assess those skills gradually until we go as far as

analyzing and critical thinking. But if they barely understand the text they are reading; you cannot go to that level of critical thinking about the text. We are giving a lot of bits and pieces from everything; again, unfortunately, we are losing some students on the way because of this.

Researcher: How is writing is assessed?

Mirjana: Writing is assessed based on a given topic, sometimes we require sentences, paragraphs, or essays depending on their level. Then they are graded. They submit their paragraphs on Turnitin, and they are graded based on those paragraphs. Then we have the writing part in the midterm and the final exam, again based on what they were taught. It could be writing a formal e-mail or like a comparison paragraph or a picture description. So based on what they have covered in the class, they are trained to do this. Whatever they were taught is assessed in the exam.

Researcher: how far are you satisfied with the writing assessment?

Mirjana: I think it should be a process. Yes, I believe it is too harsh on students to evaluate them based on one topic, one paragraph one exam. We usually give them two topics to choose from, but still, it is just too harsh on them too. To ask them to write a well-developed piece of writing in a minimal time about a topic that's given to them. This is not an external exam; this is not a standard exam in the end. This is a process. We have all students for one year. So why not have a portfolio. But again, the problem. We have different types of problems which affect everything. Number one is the levels. We have mixed-level groups. Number two is the number of students in each class. It is more than 35 sometimes. And maybe the policy imposed on us. What we should do is different. I have the feeling that we are teaching to the test. I mean, training them to produce. It is product-focused. We ask students to reach that. That is probably the end. We cannot follow up with the students because the numbers are such high numbers. Each teacher has a minimum 65/70, and they see them every single day. So, in order to follow up with them and then in order to have a portfolio work, they should be able to work very closely with the students to follow up each portfolio and to give them individual feedback to each one of them. So, it takes much time and much energy from the teachers' side.

Researcher: Maybe if you save the effort put in designing exams and grading all these kinds of things. Why? Why should it be exam-focused? I mean, that will take us to the first question again, why are we assessing them? Are you assessing them to come up at the end with a score for the skill, or are you assessing them to learn? That is was my initial question?

Mirjana: I think we should have some kind of a process. But I still believe that in the end, there must be one exam. Because, you know, we should see tangible grade. Other things are getting involved. If you have a portfolio, we do not know who wrote this paragraph. It comes with its problems, like if the student is absent on that day and then it will be difficult for him to follow up, and then they will get dismissed. So many other things are involved. So it is easier to manage an exam. You know what I mean in terms of the procedures. If

we have an exam, there is a reading, writing, grammar, vocabulary, so let us see what they do in the exam. The portfolio thing I am for it. But under these circumstances, it is challenging to administer this task. There are other things involved in that like attendance dismissals, grading, giving feedback, and students' complaints. There will be so many things coming from the part when it comes to grading and students being anxious about it.

**Researcher:** Do you believe that in the described context, it is overwhelming for teachers and students, there are lots of assessments on behalf of the quality? Maybe if we get aside all these kind of things and we focus on goals, for example, what do you think? Because there is much effort put in everything.

**Mirjana:** if you leave everything out and we exclude the assessments, then we will focus on maybe the goals. But then we will have to expose them to an exam at the end. But then this will not be fair to them because we are not training them to take an exam in the end.

**Researcher:** But then are we supposed to train them for the exam or to come up with something that they might take with them to the coming stage. I mean, this is the issue.

**Mirjana:** Yeah, but then we have to give so much importance to that one exam. Yeah. This is what we are trying to avoid, as well. And it is better to put it into pieces instead of, for example, the plan we had in mind last year was dividing the midterm into quizzes. Every two weeks, students will have bits and pieces like short quizzes, not overwhelming them with so many questions assessing minor goals. But there must be something in the process to grade them; there must be grading for their performance. It could be, as you said, it could be a portfolio, in-class assessments, quizzes, small quizzes and everything. And in the end, we can have an exam, but we cannot ignore the process. Because if we ignore this process, then, a student might just skip this process because we have students like that. And then we get to the starting point. They are not placed in the right place. They do not need to go through this process. They need to go through more advanced processes, and they deserve it. So we are just asking them to play around with these things.

**Researcher:** And when it comes to speaking.

**Mirjana:** We only have the presentation.

**Researcher:** Are you happy about them?

**Mirjana:** We should still have presentations. It is good that they learn how to present. But they should also have speaking exams.

Researcher: And what about listening? How is it assessed?

Mirjana: Listening is assessed as a quiz. It was included in the midterm and the final before. And because of the technical problems we face during the exams, we separated it from the main exams. And now we are giving it as a quiz. Again, it should be given more importance, to be honest with you.

Researcher: What kind of questions are used to assess them?

Mirjana: Multiple choice questions or fill in the gaps. We always wanted to have some note-taking skills, which is not taught or assessed. So basically, if you could change the whole system completely, then it would be better. Like it could be combined with writing so they can listen to something and write about it, but it is a significant change out of hand.

Researcher: And regarding kind of feedback learners receiving in your teaching context, whether the feedback form or type, purpose. Do they receive kind of feedback, for example, on reading assessment?

Mirjana: We assume that the teachers explain their mistakes in the class after the exams, like how they can manage it better in the next round or next.

Researcher: what about presentation or speaking, how do they receive feedback about that?

Mirjana: They have rubrics, and they are given details about the rubrics. If the students ask for the feedback because we do not want students to hear each other's feedback in the class, or if they are interested in listening to the feedback, we give in the form of scores.

Researcher: And writing. What kind of feedback?

Mirjana: Again, after the exams. Again, not in front of everybody. But if they approach the teacher and if they ask what can I do or why I scored this? So, they are given detailed feedback.

Researcher: and your viewpoint about this because last time you had something to say about tracking feedback. Yeah.

Mirjana: I think this does not work because it does not bring or take them anywhere. This task is finished. So why would I talk about the feedback? Why would I bother myself to go to the instructor and ask for feedback? Teachers do not put anything on the writing; they just write their comments; they underline certain things. We do not have this error codes any more. It should be this way. It should be a process; there should be drafts. Students should take some sort of feedback if there is a second draft, other than that there is no need for the feedback because the students know that this was tested already, I wrote it, I got my score. OK, can I change the score? Can I negotiate with the instructor? Let me try. But other than that, the purpose of the feedback session is not learning something from the feedback or improving the writing. It is just they can't change a score because they know that it is done.

Researcher: So maybe, they should be taught principles of feedback, types of feedback. The importance of feedback. I mean, things should be more transparent.

Mirjana: When I was teaching, I used to go to class after an assessment. For example, I used to take notes of the errors students made when they are writing something. And I used to go through all those mistakes without giving names. I used to prepare a PowerPoint presentation on these things. This should be done right, but it is done.

Researcher: Regarding grammar and vocabulary. How are they assessed?

Mirjana: Vocabulary and grammar are assessed with multiple-choice questions. Vocabulary is assessed only with multiple-choice questions and word formation. Grammar is assessed with multiple-choice questions plus rewrites and error recognition.

Researcher: And you happy about it?

Mirjana: I mean, I think the system has to change dramatically. Because I do not see grammar as a separate part of anything, but we teach it as a separate component, then it makes sense for me to assess it as a separate module. The problem for me is teaching it as a separate module. So, I would stop teaching grammar separately. Grammar should be integrated with other skills. It should be together with writing, speaking, it is everywhere. And vocabulary as well. We are coming up with that list, and then we are going to students and teaching that vocabulary. It should be used for ample reasons and assisting reading, but it is not accessible. So, I would change everything if we think that the coming stage would be a new joy.

Researcher: How are grading keys, rubrics, scales, schemes designed, if any?

Mirjana: I mean, yeah, we have designed the rubrics. Based on the feedback from the teachers, we made changes, but we designed the rubrics internally by the test designers. I am also involved in this. But we keep changing them. It is not something very standard or let us say this has to be followed. Sometimes, for example, the teachers want to have

more detailed rubrics, so we adapt them. We prefer to have something more global for writing, for example, holistic. But sometimes teachers want to be fairer. They want to count mistakes. They want to penalize the students. So, we said, if you wanted it that way and if it is not a burden for you. We add you know, certain things.

Researcher: What about the rubrics for the reading and the listening?

Mirjana: Since it is multiple-choice, it is just the key, and they use it. For writing, they started having the standardization meetings, and they grade one, two, three paragraphs together, sample paragraphs without the students' names to see what they give for each item in the rubric and then they tried to agree. They tried to make it standard for all teachers or graders. Presentations are also graded based on rubrics. Again, it depends on the type of presentation and

Researcher: how are issues like assessment, validity, and reliability, checked in your context are they being checked?

Mirjana: We check them internally, and we take action internally. This is our initiative. Like nobody checks the reliability and validity of the exams. We do them. We do item analysis, as we go through the questions, we see if there is a problem with any type of questions. So this is personal effort; nobody demanded that.

Researcher: how are assessment results communicated to students after the assessments are announced.

Mirjana: They are announced online, and then the papers go to the class and then students are given feedback about their papers so they can see their papers. So, it is in the form of scores, and some feedback is given in-class on the errors.

Researcher: Are they allowed to discuss the feedback with the teachers, for example? Or it is just like giving them the paper to check.

Mirjana: It depends on the teachers. It is not controlled; we do not have any control over that. So some teachers go through these questions one by one, and then they answer the questions one by one, teaching them again, sometimes revising whatever is missed, or was not answered correctly. That is how I would do. But some teachers just write the answers on the board or just give the grades. There is no adopted policy for giving feedback or communicating. It is done as per teachers.

Researcher: how do assessment fairness and ethical procedures practise in your teaching context? Are there any steps taken to ensure that assessments are fair and ethical? Or.

Mirjana: I mean, it's all personal effort. So we do not have an exam unit; we have nobody to guide us through fairness. We have an exam unit, but it is not concerned with the content of the exam. It is mostly about the procedures on the exam day. So basically, it is like administrative stuff, submission and collection. I am not happy about it. I would prefer to have someone to give us some ideas on how to ensure fairness and ethics in exams. But this is an issue or a topic that was not raised. We do it ourselves unconsciously. When I review an exam, I pay attention to. I take notes. I use four versions of the same exam with different questions but equivalent ones. I need to ensure the fairness and accuracy of this exam because they are given to different students, but I don't know how. They have to be the same in terms of the difficulty level, in terms of the length of the exams and everything, so basically we pay attention to these things, but nobody checks. Nobody asks us if they are all the same in terms of difficulty or length and individual differences. But what I do when I review one exam. I go through each question, especially for grammar, because we have many items in the grammar exam. This is an individual effort. I also try to check if this was covered in the syllabus in all classes. So again let us stress there is no guidance, no criteria, no policy, no one well-designed stuff to assure us that we are trying to do right. We just have an exam review form where we review our exams, but it is reviewed by us already, and we prepare the exams and will review the exams. It helps us to check if we covered objectives and common objectives were tested in the exam. So we just match the questions with the objectives. The checklist also highlights the duration of the exam, the chapter covered.

Researcher: But do you believe that you want to change it.

Mirjana: Yes. We have to work on it. Yeah. Maybe it would help differently.

Researcher: How far did you believe your context adopted assessment practices reflect Learner's language ability or. After the assessment and after the scores are out, do you feel they reflect their language ability because the scores are the outcome of all of the assessment that we have mentioned.

Mirjana: Not really. I think these midterms are not working. Midterms should be changed because they do not reflect what we are doing in class. They do not reflect the students' real language skills. We should also add the speaking component. Our exams are too complicated. Students find the exams too tricky for them, although the questions are not that difficult. We do much more challenging stuff in the class; they are exposed to so many different types of questions, but when it comes to exams, they do not perform that well. And that is why, I find exams not functioning as they should be. The exams are not reflecting their communicative ability and their outcome skills. For example, in this context, the students are very good at speaking. They are reasonable compared to other students in other countries, this is their strong point, and we are not testing it. They are communicative. They talk. They can listen. But again, we come to the same thing. We are trying to give them many things at the same time. We are dealing with a very mixed level of students. We are just trying to put all of them into the same box. We have restrictions. We have to follow the syllabus because we are trying to reach a certain level. So in order to reach that particular level, we have to have this bunch of things to feed the students with, because we are not only teaching general English, so we cannot follow a



task-based curriculum. Probably because we have to teach them academic skills, so it cannot be like general English. We are focusing on teaching language for a specific purpose because after all, what is the whole aim of the foundation program? This is a preparatory program, preparing them for their studies. In other words, they may need to use this language in their studies.

.

**Researcher:** Do you feel by the end of the foundation. After the assessment, so you feel that the assessments reflect the competence of the students that are needed to perform well.

**Mirjana:** No, no. Something should be done. This something should be starting with placement exams. These students should have at least three levels.

**Researcher:** Next research level you will tell us what should be done

## **Appendix 10.4: Transcription of Mirjana's Second Structured Interview**

### **(Report)**

**Researcher:** Last time, we stopped at a stage, where we talked about your adopted assessment practices and your views about them. So now we will move a step forward. In the report, you have in hand; you have what you said in both cases, for every assessment feature we discussed. Now, we want to find solutions for conflicts you raised between your contexted adopted assessment practices and your beliefs about them. If you have the upper hand to change everything or restructure everything from scratch, what would you do? What would be your assessment philosophy or purpose? Why would you do to assess language in this programme?

**Mirjana:** Okay, then I have to plan it from the beginning. Once the students are admitted to this college, they have to take a reliable placement test. Roughly this test should give us an idea about the students' level in English. Once they are placed at the right level, everything will be more individualized for them. Let us say. I would divide them into three or four levels in the first semester. If we have this placement test and if we place the students in the right level, then the elementary students will have more hours a week and more intensive course; whereas the stronger ones will have fewer instruction hours.

**Researcher:** I mean, when it comes to the assessment purpose, like, for example, what would be the reason behind your new adopted assessments?

**Mirjana:** once we have these levels, and if you are clear about these levels, then we will have objectives for each level set clear from the beginning. After we set the objectives, it will be easy for us to check the progress through the assessments, exams, assignments, and project work. So the assessment would be for checking the progress of students to reach a specific objective within their level. So it is like you link assessment purpose to course objectives. So we have to deal with these objectives bit by bit. However, we cannot deal with all of them together with all students. If the students achieve those objectives, then they will move on to the next level with new objectives and with new assessments, new books, and materials depending on the curriculum. This will be like a smooth transition from one level to another level, and then it will also give the students the chance to repeat things they haven't achieved, but not the whole year or the whole semester. So they will have to repeat maybe, six weeks of the semester, and then they will continue from where they were left with; they still have the chance to catch up.

**Researcher:** And you would link these objectives to the goals of the program.

**Mirjana:** Yes. The goal of the program is like to reach the BE2 level. The B2 level is the level mentioned in the common European framework and accepted by many institutions. It should be like an intermediate level in general English supported by academic English as well because we have to prepare them somehow for academic English as well.

Researcher: Would it be better if you said like, for example, instead of having this as a general English B2 level, you would be linking this to the studies they are going to go through? Because this whole programme is a preparatory program for academic studies that are run in English with a certain level.

Mirjana: Yeah, but we have to take them to that level first.

Researcher: What about conducting needs analysis?

Mirjana: If we bring them to that level, they will be prepared for any type of course at the university level. We are making it only to English100 run in liberal arts. We are not trying to make them understand English 100 or be successful in English 100; we want them to understand everything given to them in English because the language of instruction here in this college is English. So they will study math in English, engineering in English. So they should have enough English to understand their studies. That is why we cannot only focus on academic knowledge; we should focus on more general academic knowledge (English for general academic purposes) not on English for specific purposes. This is an initial step; maybe further specific ones could be done later on in undergraduate program depending on the type of programs they would be studying because those who are studying business are different from those who are studying engineering and the requirement would be different. These common European framework objectives are set objectives for each level. It starts with A1, and it goes up to C2. So basically, for each module, they have set goals, specific things for students to do, so it is measurable. Based on these objectives, we set our programme objectives, and if we reach this level that is required to study in an academic environment, then we are okay.

Researcher: So it's not an assessment of achievement. It is an assessment of achieving objectives.

Mirjana: It is not about one course or one program; it is about their language competency, which they need for their academic studies. It is not a general English course. It has to be related to academic. When we choose the book, we cannot just choose it randomly. We have to choose a book suitable for their studies. This is the second step; we will change the curriculum. Once we set the goals and objectives, we have to find materials that would meet these objectives. If they do not meet our objectives, we have to supplement with our material, and then we start bit by bit.

Researcher: Regarding assessment sources or preparations. What would you do? Would you take courses? Would you study? How would you do it?

Mirjana: I think whoever is designing exams, assessments, or projects should be trained. I would go for some type of training sessions, which would be given by experts. Also, before we give them the task, we should have some workshops, and we should sit down and discuss certain things and check what we have learned from those workshops shops

and how we can apply what we learnt in our context. Everybody needs to perceive some kind of training. This is the first step. Before we start doing the fall and spring semester, we have to finalize everything in the summer, such as training, syllabus, levels, placement exams, and how they will be administered. So, everything should be ready. I would ask for the assistance of well-qualified people to assist me. They should all be trained before doing so.

**Researcher:** And what about assessment source? Would you adopt externally mandated assessment, or would you design your assessment?

**Mirjana:** For the placement exam, I would adopt an external source because they have those general standardized exams for placement to check the students' level because I am not trained to prepare assessments like that. So basically, we can leave that job to them, and we can adapt their tests.

**Researcher:** And then for your classroom assessment?

**Mirjana:** We have to use our resources, questions. We can use books as resources.

**Researcher:** What about other teachers who are not involved and not taking any assessments responsibilities?

**Mirjana:** I would form a committee with some experienced instructors, but they have to be trained because people do not know how to create or questions and assessments in general; otherwise, we would face conflicts, things might not be accepted. At least people will understand why we accept or refuse something; they will not take it personally.

**Researcher:** What about peer or self-assessment. How would you involve it in your new program? Would you implement it? Would you put it as one of your assessment criteria?

**Mirjana:** The thing is, we should change the type of assessments. I am not for setting mid-term or final exams, especially for writing and speaking. I think they should be tested differently. Writing should be assessed through a portfolio. We should adopt process writing so that students can be involved in peer review and self-review. If there is no option for improvement, no room for improvement, so why would they review that?

**Researcher:** When you come to implement something like this, I mean, as we mentioned, you need to have the background about principals or framework for how to do, how you would implement without having this kind of experience, from where you would get the source about how to implement peer and self-assessment. How do you prepare for the step?

Mirjana: I have not tried it before officially, I am doing intuitively in class. We asked students to grade each other, but then these grades were never used. I do not know if I could give it some weight, but I would indeed use peer assessment or self-review for their developments, not grading. I would give it as a formative assessment like I would provide them with checklists or rubrics, but not necessarily giving them the grades, just for improvement reasons.

Researcher: What would be the source of your techniques?

Mirjana: We have to train the students to use this, how to use the rubrics to grade each other's work, use error codes and things like that sort. So, we have to train the students as well if you want to involve them in this process.

Researcher: What about assessments of different skills. How would you assess them?

Mirjana: We have common assessment problems in this context. I think we have to tackle those problems very closely. We cannot ignore them because it does not work. We have to take everything step by step. For writing, mostly, it is their weakest point. They have to start with sentence structure; it should be taken step by step until we reach essay writing in the second semester. If we cannot reach essay writing, it is okay because, in the undergraduate programmes, they work on essay writing. However, we cannot ignore that they cannot use punctuation correctly. They have many fragments. They have many run-ons and long sentences. So we have to deal with these things first from the beginning. The root for writing is writing sentences, so we cannot build on something, which is collapsing already. If we have levels, especially the first level should work on sentence structure, we have more time to spend on each level.

Researcher: You mentioned something related to process tracking. How would you implement that?

Mirjana: Yeah. If you start teaching as a process, then we will test writing as a process as well. If they are working on sentence structure, so maybe they will focus on writing one or two sentences at the beginning for the first assessment, and then, later on, they will increase the number of sentences.

Researcher: How would you assess the sentences?

Mirjana: I would use rubrics, of course. However, again, we have to have elementary rubrics for sentence structure. We cannot have many components for the rubrics. We can use added error codes and give them feedback. You have to change it that way so they can have another chance when it comes to their second assessment and the third assessment. Let us say at the end of the term; they will have a folder which includes

their work from day one so that it builds on each other. And then maybe it will become a story at the end. Maybe it will become a paragraph. Maybe it will become an essay at the end of the year, but each individual should be having or keeping a portfolio about his/her writing to monitor their improvement. It makes a huge difference.

Researcher: And the kinds of error codes you mentioned, would you count on error codes?

Mirjana: No, because we used to teach error codes, and we used to teach them every course. They have to be taught in the class, and they might not mean anything to some of them. Some students might need some kind of feedback, like face to face feedback, interviews, interaction with the instructor. That is why we have these writing centres, labs in colleges and universities. So if the instructor cannot give that kind of feedback, those students should seek help from other resources on campus. This should be given much weight, and the students should know they have to improve their writing; it is their responsibility to improve themselves, which is self-assessment. This is what is missing here because we only evaluate one thing at a time. But we do not evaluate the process. Students are not aware if they are progressing or not. So, they cannot assess themselves in this system. If we want them to assess themselves, they have to see the whole process. They should see the process and improvement so that they can assess themselves or they can assess another friend. Otherwise, writing paragraphs will be written exams. Even if I do give feedback about that paragraph, where are you going to use that feedback? You are not going even to read it because it is a past exam.

Researcher: What about reading and listening?

Mirjana: We need to focus on reading and listening skills. These are respective skills, so basically the students are supposed to read or listen to something for sure, and they do it. They do it automatically; it does not have to be like a text. It does not have to be a recording all the time. When we talk to each other, we listen to each other. However, for formal assessment, different types of assessment will be created again based on their level. I cannot ask them to deal with two-page if they are on the first level. If they are dealing with only small paragraphs in their classes, we cannot ask them to do more in their exams. We can have small quizzes like here and there to check different types of reading questions, testing different objectives in reading and different skills in reading so that they are familiar with what they are supposed to do because most students fear reading tests. They think that they will not understand the text. So, in order to break the fear, we have to come up with something, like in-class assignments that focus on skimming or scanning. They can learn to focus more on reading skills.

Researcher: And the type of questions and the item, what kind of questions would you add?

Mirjana: Matching open-ended questions, multiple-choice, insertion, filling in tables, writing a paragraph based on something, comparing two characters. We need to do something with reading. Maybe a task-based one because it involves authentic

procedures. In real life, we read for a reason, for learning something, to do something with this reading. We should give them a reason for reading or listening. I would add newspaper reading, some notes from the social media or Twitter, some messages on Twitter because they read these things. Then we can ask them to write about what they think, what they understand, and what this person means, what is the main idea behind this message. So we could assess reading by asking them to speak, create argument maybe or like synthesizing the information. We can give them, like two texts with opposite arguments, and then they can come up with their arguments. Such things would work better, of course. However, the thing is, when it comes to testing, it is difficult to assess these things. So maybe we need to move away from this testing culture to an assessment culture, from being test-oriented, grade-oriented, and score-oriented to being performance-oriented. However, at the same time, the problem there starts with fairness; when we move a little bit away from the tests and when we move on to the assignments, tasks, and assessments, it is difficult for some teachers to receive feedback from us. They do not accept feedback themselves, and they do not want to give any bad feedback to the students or any constructive feedback to the students, they want to play "Everything is perfect". It is a teaching culture. There are lots to be changed in order to implement what you are convinced of. Once you start implementing such a vastly different change, then you will have to treat everybody very well. You have to believe in them like you have to trust them. You have to trust the stakeholders in this new system. Do you know what I mean? Like, you have to trust the teachers. You have to trust the students. They should all have the same objective every day. Some teachers themselves are test-oriented or students-oriented. Moreover, in order to implement those massive changes, it should start with transparency with all its supporters. It is not easy. If this is the ideal world, it would be great, If it were my decision and if I could manage it, this is my dream. This is how it should be like; we should not focus on the grades. Students should be very relaxed. We should focus on their learning issue, achieve our objectives in the end, so our focus should be on complete change. Nevertheless, at the same time, in reality, when I think about it, when I consider other things related to this, it is not easy. That is why I think in the world, all around the world, at some point, people still depend on exams because when the human factor is involved, things will not be objective. Unfortunately, at some point, things will change, things will be different, some people will have hard feelings against other people, and it will affect their performance, it will affect their fairness; some kind of human factor is involved so that emotions will be involved as well. It is not like an IELTS examiner, who is welcoming students whom she does not know at all. So, she's fair to everyone as she has no issues, no relationships, no one. So, her fairness degree would be better than the person who has some kind of interaction with the system throughout the year. Do you know what I mean?

Researcher: This would take us to grammar and vocabulary

Mirjana: They should not be tested separately; they should be tested with the other skills, with writing and speaking. I do not say that they are not necessary, of course. They are essential parts of the language, but what should be tested is their usage, especially if we are talking about academic writing.

Researcher: And if we're talking about assessment design, the one that we talked about, how would you implement it? How would you change it? It is like we mentioned

something like setting purpose, writing the specification, starting with some design and then testing the design, validity of items. How would you do this?

Mirjana: If you are talking about a radical change, like not having exams and having some assessments and tasks, then we do not need to make that. We will not have any items to be checked. Do you know what I mean? It will not be like, tests. We will do all of these assessment types throughout the whole semester, but in the end, there should be some benchmark, as a proficiency exam. It could be an external or an internal one. If we leave the midterms and finals behind and if we adopt a new system of assessments, then we will not need any of these. We will focus on the performance of the students.

Researcher: And for designing rubrics and answer keys, what would you do about them. You mentioned something about Turnitin, how would you use it efficiently in your system.

Mirjana: students should submit everything to Turnitin. So that they can see them all the time, they can create their folder, soft folder as well not only hard copies. They can also have soft copies' they can receive the feedback on online. They can see the feedback online, and they can work on it even from their phones. So I think it would be more practical if you use the codes and if we give the feedback online. Technology should be involved, maybe video conferences, maybe why not? Because you can also leave feedback and audio feedback on Turnitin.

Researcher: What about administration stuff?

Mirjana: We will cancel all of this if we are not going to be using final and mid-term exams. Any we will base all our program on classroom assessments.

Researcher: What about validity, reliability?

Mirjana: We have to do some regular checks even if they are classroom assessments and this test, in the end, should be there. We have to check if the students really achieved those objectives or if the teachers also push them a little bit, you know what I mean. We have to have something standard at the end because our primary goal is to make students reach that level of language. Right. So we are trying to make all students reach the same level. Somebody may get 90. Somebody may get 65 or 70, whatever. However, we want them to reach that stage at that level. So we have to check if they reach those levels in order to give them the pass. The process can be more flexible, more in-class assignments, more based on development and learning. But at the same time, in the end, we have to assess their achievement.

Researcher: How would you manage anxiety either from students' part or teachers' part. You are starting with this formal standardized or external assessment in order to divide them into their levels, and then you start the ongoing assessment in the classroom, and



then you go back to the exam at the end. How would you train them for the exam? How would you specify what will come to the exam?

**Mirjana:** They can be trained. Of course, once they reach this last level before the exam is taken, they have to be trained. They have to know the types of questions, and what kind of questions are expected in these exams. I mean, it could be similar to an IELTS exam or a TOEFL exam. It could involve some kind of speaking in it, why not? So mainly teaching and an ongoing assessment that would make us sure that we achieved our objectives and at the last stage we can start another process in the last month or something like this, preparing for a standardized assessment especially we have three semesters. They can attend the summer school for preparation for the proficiency exam, and at the end of the summer school, they take the proficiency exam. Do you know what I mean? If they fail the proficiency exam, they will not repeat the whole course from scratch again. They will only get the summer school for four weeks, three weeks, or two weeks just for the exam training.

**Researcher:** How would you use the results from the exam? How would you communicate them to students? What would you do in order to make the students benefit from such an experience? For example, how would you analyze the weakness and strength?

**Mirjana:** The results will give us an idea about what we should do next. We have to do item analysis for our feedback. But for the students, again, if they passed, it is good because this is the only test they have or they are supposed to take but if they do not pass, they need to go through this training like for two weeks.

**Researcher:** Before you do the training, you need to know. Maybe do some sort of analysis on the results before you communicate them with the students, to the audience maybe you tell them about their problems.

**Mirjana:** We can have individual meetings with them based on their exam results, things they missed, things they should work on, and areas they are weak at. They have to know their pitfalls, their problems because we are expecting very few failing students. Suppose you go by this system. In case they do, we will have those sessions may be for two weeks, three and then again have to revisit the exam. However, it is not like repeating the course because we would not let them sit for the exam unless we are quite sure that they acquired what is needed; it is a matter of exam training.

**Researcher:** What about fairness and ethical procedures?

**Mirjana:** That is the question that crossed my mind. If the assessment is subjective and if it is based on learning and achieving objectives only without any numbering, scoring, rubrics, and questions, what would happen? If you leave that idea behind and start something completely, then we could be under question.

Researcher: But one of the aspects of fairness is to give people what they deserve or to allocate them where they are. You went for ethicality at the very beginning. You trained them. You satisfied your conscience by making sure that they acquired everything. So actually, you maintained ethicality and fairness throughout and again when it came to tests, you trained them. You made them sit for the test.

Mirjana: Still, I am not convinced.

Researcher: How would you assess?

Mirjana: Because the problem is that when it comes to subjective evaluation, students can argue even if they get 90 for their writing, why not 95.

Researcher: one of the principles of critical language testing is to involve students in assessment, not only in the sense of peer or, self-assessment but also in assessment decision making.

Mirjana: If we make such a radical change, then students have to grade themselves first, maybe if they reach fairness themselves, then we can reach fairness of the exams.

Researcher: What about asking about their own opinion about what you were doing? Maybe their opinion about your courses would help you to modify and change, and this is part of ethicality.

Mirjana: The problem is that it is not easy to break. Students have come with such background where the main idea and focus is the score and the final grade. Now it is difficult to break this habit of focusing on grades and focus on evaluating their self-performance and self-learning and convincing them that this is for their development. This is for their learning.

Researcher: But if you involve them in decision making from the very beginning, maybe they would help you.

Mirjana: You know, everything should be said in advance. As I said, everything should be finalized before the semester starts.

Researcher: How would you evaluate such kind of experience that you are sharing? I mean, do you believe in everything you've said? Do you want really to implement?

Mirjana: I would love to implement it.

Researcher: What are your challenges?

Mirjana: Fairness would be my challenge, the biggest challenge. I would not be sure to eliminate the threats on behalf of objectivity for assessment. This scares me, and I do not know how to do it. Teachers' opinion would be one big obstacle because based on my experience so far in this context, it is challenging to reach objectivity. When it comes to teachers, when it comes to students, when it comes to real satisfaction, it is difficult to satisfy the students. It is difficult to satisfy the teachers. It is difficult to satisfy the administration because everybody is afraid. That is why it is so difficult for teachers to be fair or to be objective, and they prefer to be lenient. I'm not saying that they should be stricter and they should penalize students for every mistake. The fundamental change is against this idea of penalizing students.

Researcher: What is the chance of implementing something like this?

Mirjana: I think the only way to implement such a thing would be still having some kind of formal exams going together with the main ones. I think the placement and the exit exam should be better for the program. However, during the process, the portfolio is essential; students should be graded based on in-class assignment; midterms and finals should be left out. Maybe there must be some kind of paper documentation for grades to defend yourself in front of students, teachers, and the system. Once the teacher gives the grades, she needs to base it on something to defend herself and to justify her grades. There should be an explanation for everything. And maybe it will take time in order to write these, so the feedback sessions should be recorded. Maybe feedback should be audio or video recorded in an online tutorial. There must be some something to ensure trust. Those are students, and there must be some kind of control.

I would be satisfied with this new agreement. It will be a significant change. We will see what worked and what did not work well, and we could make changes accordingly. I would love to implement such a thing. But I would be open to feedback and open to mistakes and errors that would come across that because you can never know what kind of challenges are waiting for us there. And we can anticipate some of the problems or challenges like I said, fairness, but there will be many more. If you were to make such a significant change, we have to stay determined.

Researcher: What about development. This will require continuous development for teachers, students, for yourself, academic development, as you said, training.

Mirjana: Teachers should be trained on giving feedback, grading written assignments, grading speaking assignment, being objective, being fair, and preparing assignments because now it will be more individualized. Do you know what I mean? Teachers should have some kind of training about delivering the messages, instructions, exams, and feedback to the students.

Researcher: you admitted that you could not stop being anxious about the new changes but at the same time satisfied.

Mirjana: I would feel more satisfied compared to the system now. I would feel motivated to work on it because I believe it would change, and it would improve the whole thing. The whole system; but it is not easy.

Researcher: if you will give one word/phrase to describe the current system and one word/phrase to describe your dream situation.

Mirjana; Now, it would be too harsh to say Chios. Then if I had this dream system, so many good adjectives would come to my mind, but I would say: Student-Cantered, Learning-Oriented, and More Democratic More Open-Minded. Very Satisfactory from my side, And Very Organized if I compare it to the chaos happening now because we will start the whole thing in a very organized way.

### **Appendix 10.5: Transcription of Mirjana's Completed Report**

<b>Assessment Component</b>	<b>Current Practice</b>	<b>Participants' Belief</b>	<b>Solution</b>
Assessment Purpose	<p>We are trying to reach roughly B2 level based on the Common European Framework, which is enough for understanding and studying in the English language so they can continue their studies in their academic majors without difficulties by the end of the program. We assess to see if they can do the requirements of that level. Graded summative assessments throughout the academic year evaluate the achievements of these goals and provide grades. In the end, an accumulative grade is given, and a student would be considered as pass or fail according to a pre-set total percentage (70%). The reason behind assessments is to see if they have reached that level. We are assessing the achievement of specific goals or objectives.</p>	<p>The adopted assessment purpose is for judging the achievement of specific goals or pre-set learning outcomes. Some assessments are done throughout the semester to check to learn even though they are not official graded ones, and they do not provide official feedback. If you ask me I believe in more formative assessments than in summative ones so that we can see the process and where learners are so, we can take action accordingly instead of waiting until the midterm exams, which is in the mid-semester and it could be too late for some students to catch up after that period. The problem is that formal assessments are still dominating because their percentages are high.</p>	<p>Once we have these levels, then we will have objectives for each level set clear, so assessments' purpose would be for checking achievement of each level's objectives; if students achieve those objectives and show related learning outcomes in their assessments, then they will move on to the next level with new objectives, material, assessments, depending on the curriculum. This will be a smooth transition from one level to another level. So it is not an assessment of achievement anymore. It is an assessment of the achievement of pre-set objectives. I would feel more satisfied compared to the system now with this new one I would feel motivated to work on it. Because I believe it would change and it would improve the whole thing. The whole system; but it is not easy.</p>

Assessment Source	We are doing them internally.	I am with internal mandated assessment because we have to have assessments that suit our context and our learners' need. We cannot rely on external resources to assess our students with their different levels, backgrounds, language skills, and unique specific culture. So, I think the more individualized assessment sounds, the better they are because it is a sensitive context.	For the placement exam, I would adopt an external source because we need those general standardized exams for placement to place them into levels. And then for our classroom assessment, we have to prepare them and use our resources and design our tasks. We can use books as resources. If we leave the midterms and finals behind and if we adopt a new system of assessments, then we will not need any of these. We will focus on the performance of the students. However, when the human factor is involved, things will not be objective. In the end, there should be some benchmark as a proficiency exam. It could be an external or an
Assessment Methods	We have formal graded assessments in the form of mid-term and final exams, where we assess reading, writing, listening, grammar and vocabulary. We have less formal assessment like tests, quizzes, assignments, projects, individual and group presentations students are basically in their comfort zone, in their classrooms. These are graded, but they take place in a very relaxed	I am satisfied with the Variety in assessment methods we use, but I believe we need to manipulate them in order to make them maybe more authentic to reflect the real competence of students. I am not that satisfied with the presentation as the only tool for assessing speaking skill. The problem that students prepare them and present what they prepared; it is not an	We should change the type of assessments. I am not for setting mid-term or final exams, especially for writing and speaking. Learners should be assessed in different ways: task-based assignments, projects, short quizzes. By the end of the course, once they reach this last level before the exam is taken, they have to be trained. They have to know the types of questions. What kind of questions are expected in these exams? I mean, it could be

	<p>atmosphere, so students do not feel anxious about them. They do not have very high percentages, but it is good to make them produce something based on the language they acquired. I think they serve very well in terms of seeing the weaknesses of the students and work on them before the exams.</p>	<p>authentic performance. I think there must be a speaking exam, where students are given a topic to talk about, and teachers would assess their accuracy and fluency. Some grades are awarded to attendance, which I do not consider it an assessment method; I would not penalize or award the student for attending a class. Sometimes the assessment methods or tools are not used the way they are meant to; students are graded, for example, in projects on bringing visuals.</p>	<p>similar to an IELTS exam or a TOEFL exam. So mainly teaching and an ongoing assessment that would make sure that we achieved our objectives and at the last stage we can start another process in the last month or something like this, preparing for a standardized assessment.</p>
Assessment Decision Making	<p>They are the team leaders and the head of the department. These people prepare the exams, and the head of the department reviews the exams, and they are stored and printed without having the instructors look at it.</p>	<p>They are not entirely excluded from the process because they know the format of things and the types of questions like they are given the instructions to mark assessments. I think it would be a better idea to involve them more in the preparation and designing processes because they teach the course and know their students well. Team leaders also teach, but they are only exposed to one group of students. If</p>	<p>I would form a committee with some experienced instructors, but they have to be trained because people are not trained enough to perform this assessment role. We would go for some training sessions, which would be given by experts in assessments. Before we start the assessment task, we should have some workshops for practical training; then we should sit and discuss how to apply what we have learned in our context. We cannot just take everything and use them as they are. We</p>

		<p>teachers are involved, there will be more discussions; assessments will be more individualized and more customized based on the needs of students. Plus, burden and responsibility will be shared, and the load on team leaders will be less.</p>	<p>have to adapt it to our context. Other teachers may get involved by sharing ideas, suggestions but after receiving training from the committee. At least people will understand why we accept and or refuse their ideas; they will not take it personally. Whoever is involved in assessments should be trained. Teachers should be trained about giving feedback, grading written assignments, and grading speaking assignments. Being objective, being fair, and preparing assignments, because now it will be more individualized.</p>
Assessment Design Process	<p>There is no specific process. We have a checklist that we need to fill regarding duration, covered chapters, number of versions, learning outcomes we are assessing, and pre-set types of questions and items that we have to follow strictly with specific percentages regardless of the course nature or our opinion.</p>	<p>I am not happy. We do it intuitively. We are not guided; there are no specific criteria that we all follow. We need guidance, training and validation for what we do. It is a huge responsibility.</p>	<p>If you are talking about a radical change, like not having exams and having some task-based assessments, then we do not need to make that. It will not be, test-based. We will do different formative assessments throughout the whole semester.</p>
Assessment of Language Skills	<p><b>Reading</b> skill is assessed in mid-term and final exams and quizzes in the form of reading texts followed by multiple-choice and true/false questions. <b>Writing</b> is assessed in mid-term and final exams, where students are required to</p>	<p><b>For reading</b>, I believe the type of questions should be modified; it should only focus on open-ended questions. Unfortunately, the current structure is rigid and enforced on us. We need to focus on their competence in</p>	<p>For writing, it should be monitored through a portfolio. It should be process writing. We can start with sentence structure because we need to take them step by step until maybe essay writing in the second semester. However, we cannot ignore that they cannot use punctuation</p>



	<p>write sentences, paragraphs, or essays; besides, it is assessed as part of the project, when they submit their paragraphs as group submission in Turnitin for grading. The writing topics are related to topics they were taught about in class. <b>Speaking</b> is assessed by asking students to prepare presentations and present in class. <b>Listening</b> is assessed as a quiz. It was included in the midterm and the final before, but because of technical problems, we separated it from the main exams. Moreover, now we are giving it as a quiz as a listening text followed by multiple-choice, true/false or fill in the gaps questions. <b>Vocabulary</b> is assessed only with multiple-choice questions and word formation. <b>Grammar</b> is assessed with multiple-choice questions plus rewrite and error recognition.</p>	<p>comprehending the text. We can use matching ideas with paragraph. I will change the questions in every exam based on the text itself. We are teaching many reading skills, but we are not assessing them like referencing, insertion, skimming, scanning, summarizing, analyzing. There are many skills that we do not assess like critical evaluation, analyzing, re-ordering. <b>In writing</b>, I believe it is too harsh on students to evaluate them based on one topic, one paragraph in one exam and to ask them to write a well-developed piece of writing in a minimal time. It is product-focused practice; we are teaching for the test. It should be a process, where we use portfolios for example, but we will face three challenges: mixed level of students in the same class, number of students, time, and most important the context adopted policy. Plus it will come with its problems: absence, interest in grades from students' side, grading, giving feedback; so maybe</p>	<p>correctly. They have many fragments. They have many run-ons, long sentences. So we have to deal with these things first from the beginning. The root for writing sentences, so we cannot build on something, which is collapsing already. If we have levels, especially the first level should work on sentence structure, we have more time to spend on each level. Start teaching as a process, teaching writing like a process. Then we will test writing as a process as well. For reading and listening: I would work on them together; we can have small quizzes like to check different reading and listening so that they are familiar with what they are supposed to do because they fear long texts. We can vary the type of items that assess those skills. We can ask them to perform authentic tasks based on what they read or listened to similar to real-life ones. They have to do something with the information they extract from any visual or audio text.</p> <p>We could vary the genre and registrar of texts they are exposed to. We can use maps, newspaper, and messages on social media, and the kind of required tasks will vary accordingly. We can ask them to</p>
--	--	---	---

		<p>we can use with an exam at the end. <b>For speaking</b>; we should not depend only on presentations; we should have a speaking exam. <b>For listening</b>, we always wanted to have some note-taking skills, which is not taught or assessed. It could be combined with writing so they can listen to something and write about it, but it is a significant change out of hand. <b>For grammar</b>, the system has to change dramatically; I do not see grammar as a separate part. The problem for me is teaching it as a separate module. So I would stop teaching grammar separately. Grammar should be integrated with other skills. <b>For vocabulary</b>, it should be used for ample reasons and assisting reading.</p>	<p>write or speak about them, so we are assessing their language use. They can create an argument through writing or speaking. We need to move away from this testing culture to an assessment culture, from being test-oriented, grade-oriented, and score-oriented to being performance-oriented. For grammar and vocabulary: They should not be tested separately, they should be tested with the other skills. I do not say that they are not necessary, of course. They are the essential parts of the language. However, what we should assess how they are used.</p>
Assessment Administration	<p>The exam unit does this one. We are involved in printing. We submit, and they take care of storage, and collection on the day of the exam.</p>	<p>I do not believe we should be involved in administrative stuff like exams printing.</p>	<p>We will cancel all of this if we are not going to be using final and mid-term exams. Any we will base all our program on classroom assessments.</p>
Assessment Grading	<p>We design rubrics and answer keys internally for open-ended and closed-ended assessments,</p>	<p>I think we need to have standardization meetings, where we grade samples of paragraphs together</p>	<p>We are now talking about grading tasks, not exams and tests. We have to have simple rubrics with no many overwhelming items;</p>

	<p>respectively. It is not something very standard. We prefer to have something more holistic. Teachers sometimes want to have more detailed rubrics, so we adapt them. They want to count mistakes and penalize students.</p>	<p>without students' names to agree on specific criteria that can work based on what we have.</p>	<p>teachers will be trained on how to design them in advance. Teachers can use error codes and feedback during grading. Rubrics should be uploaded inline, where learners are submitting their tasks.</p>
Peer or Self-Assessment	<p>We do not have them formally. Some teachers would provide learners with writing or speaking rubrics and ask them to grade oneself or others.</p>	<p>In my opinion, it should be part of the assessment. Like, for example, in writing assessments, we have checklists for them to check after they finish writing. They are not grading themselves, but they are checking if they have completed tasks. So it is kind of a self-assessment as well. Nevertheless, again, students should be trained; they should know how to look at certain things in their performance. Some learners are harsh on their peers because they are not trained. When it comes to self-assessment, they are influenced by their background because their main concern in this context is the grades. Even if they know that they do not deserve the grade, they still want to have the grade. Once it comes to assessing</p>	<p>I would instead use this peer assessment or self-review for their developments, not for grades. I would give it as a formative assessment, like checklists or using the rubrics on their work or their peers work, but not necessarily giving them the grades. We have to train students to use rubrics to assess their work and others or use error codes. We have to train the students as well if you want to involve them in this process.</p>

		<p>themselves, they will all give full points to themselves. They will not be critical. They do not reflect on themselves; they do not have the skill. Teachers/assessors should also have the skills, tools and techniques of doing it.</p>	
<p>Assessment Feedback</p>	<p>After grading exams, teachers take them to class. We assume teachers explain how to manage them in a better way the following time. For writing and speaking, we do not discuss rubrics with them in class so as not to embarrass them; we give it in the form of a grade. If they ask for feedback, they can be given once in the teachers' offices.</p>	<p>I think this does not work because it does not bring or take them anywhere. Teachers do not put anything on the writing; they underline certain things. We do not have this error codes any more. It should be a process; there should be drafts. Students should take some sort of feedback if there is a second draft, other than that there is no need for the feedback because the students know that this was tested and what they will do is negotiating grades. The purpose of feedback should be learning something from for improvement.</p>	<p>Students should submit their tasks to Turnitin. So that they can see them all the time, they can create their folder, soft folder as well not only hard copies. They can also have soft copies; they can receive feedback on online, and they can work on it. So I think it would be more practical if you use error codes that we train students on comprehending them in addition to descriptive feedback online. So. Technology should be involved, we can use video conferences, and you can also leave audio feedback on Turnitin. Some students might need some kind of additional feedback, like face to face feedback, interviews, interaction with the instructor. That is why we have these writing centres, labs in colleges and universities. Those students should seek help from other resources on campus.</p>
<p>Assessment Result Communication</p>	<p>They are announced online, and then the papers go to classes where students can</p>	<p>I am not satisfied with it. There should be an adopted policy for giving</p>	<p>The results we get from the ongoing assessment will feed the system, in other words, we will</p>

	<p>see them. So it is in the form of scores. As for feedback, there is no system for that. It is not controlled; we do not have any control over that. It is done as per teacher.</p>	<p>feedback or communicating it.</p>	<p>use it to adjust teaching, develop learning and learners so that they can achieve their intended objective to be able to go to the following level. So we will analyze their outcome of the ongoing assessment, give feedback, follow up with learners, reassess the objectives. We would not let them sit for the exam unless we are quite sure that they acquired what is needed. Results from the proficiency tests will be analyzed to know the strength and the weakness of each learner; they need to go through more training based on their weak areas only until they reach the intended level, they do not need to repeat everything all over again. We can have individual meetings with them based on their exam results, and what questions they missed, what questions they should work on, what areas they are weak at. They have to know their pitfalls, their problems because we are expecting very few failing students.</p>
<p>Assessment Reliability</p>	<p>It is not checked. We analyze item difficulty only intuitively; we go through the questions; we see if there is a problem with any type. So this is personal effort.</p>	<p>I am not satisfied because we do not pay attention to these concepts. Assessors should be trained to perform this task.</p>	<p>If you were to adopt such a system, we have to do some regular checks even if they are classroom assessments and this test, in the end, should be there. We have a reliable tool to check if the students achieved those</p>

			objectives. We have to have something standard in the end because our primary goal makes students reach that level of language. So we have to check if they reached that level in order to give them the pass. In the end, we have to assess their achievement.
Assessment Validity	It is not checked	Higher management and supervisors should check it and be trained on how to verify it. Assessors' work should not be left to intuitions.	If you were to adopt such a system, we have to do some regular checks even if they are classroom assessments and this test, in the end, should be there. We have a reliable tool to check if the students achieved those objectives. We have to have something standard in the end because our primary goal makes students reach that level of language. So we have to check if they reached that level in order to give them the pass. In the end, we have to assess their achievement.
Assessment Fairness & Ethicality	We do not have an exam unit that guides us through fairness. We have exam unit, but it is not about the content of the exam. It is mostly about administrative submission and collection. We do checking ourselves unconsciously especially that we create versions of exams. When I review an exam, I pay attention to	I am not happy about it. I would have someone to give us some ideas on how to ensure fairness and ethics in exams. However, this is an issue or a topic that was not raised. So again let us stress there is no guidance, no criteria, no policy, nor one well-designed stuff to assure us	One of the aspects of fairness is to give people what they deserve or to allocate them where they are. You went for ethicality at the very beginning. You trained them. You satisfied your conscience by making sure that they acquired everything. You made them involved by assessing themselves and others; so actually, you maintained ethicality and

	<p>difficulty level, length of the exam, alignment with objectives, but nobody checks.</p>	<p>that we are trying to do right.</p>	<p>fairness throughout. However, the question that crossed my mind. If the assessment is subjective and if it is based on learning and achieving objectives only without any numbering, scoring and rubrics, then we could be under question; that is why I came up with the idea of the proficiency test at the end, we train them, we make them sit for the test.</p> <p>Students have come with such background where the main idea and focus is the score and the final grade. So now it is difficult to break this habit of focusing on grades and focus on evaluating; there is a pressure going on here. Teachers' opinion would be one big obstacle because based on my experience so far in this context, it is challenging to reach objectivity. When it comes to teachers, when it comes to students, when it comes to real satisfaction, it is difficult to satisfy the students. It is difficult to satisfy the teacher. It is difficult to satisfy the administration because everybody is afraid. Thus, that is why it is so difficult for teachers to be fair or to be objective, and they prefer to be lenient. I am not saying that they should be stricter and they should penalize the students for</p>
--	--	--	---

			<p>every mistake. The fundamental change is against this idea of penalizing students. I think maybe the only way to implement such a significant change would be still having some kind of formal exams going together with the main ones. I think the placement and the exit exam should be better for the program. There must be paper documentation of grades to defend yourself in front of the students, teachers, and the system.</p>
--	--	--	---



## **Appendix 10.6: Transcription of Mirjana's Unstructured Interview**

Researcher: This is the last phase of this study. It is an unstructured interview. The first part will talk about your experience with the different phases of the current research. How do you evaluate it, or how do you find it?

Mirjana: It was a kind of reflection about the phases I went through throughout my teaching years. I have been working as a teacher, assessor, team leader, head for almost 15 years now. So it gave me a good reflection on what I have been doing and what could be done in a better way and then what works fine. It is good to revise all these things from time to time. It helped me see that some things are not necessary, whereas other things are necessary and they have to be there. They have to be given more importance. For example, I realised that the assessment process should have more involvement from the students' and the teachers' side. I am happy to be involved in that because I had an excellent experience thinking about what we do here and being critical about the things that we are doing in this context and how we can improve things in terms of assessments. So it might give me the chance to propose some changes in the future to improve the assessment mechanism here in this context.

Researcher: Regarding the checklist phase, how do you evaluate this experience?

Mirjana: I have realised that I do not know much about the theoretical part of assessments, whereas I do implement things in practice. I felt the need for more training and support. I would appreciate some support from the management in terms of assessment, preparation, assessment, design and evaluation process in general, but at the same time, I also felt comfortable that we do some things in there in the right way. I also realised that not all theoretical things matter. So basically, practice is more important than theory, as usual, and some things are not even needed even if they are given importance in specific contexts. For example, I always thought that statistical data are needed. However, when I think about it now, I feel we do not need much detail about the results of the exams or the averages or the statistical data. Yes, we have to have some feedback about an item analysis to give us an idea about the exam questions, but I think the more important part is that we should change the whole system, and then we will not need the statistical data at all because we will not depend only on exams. Instead, we need to assess the students' performance and learning in different ways. I think the main idea should be giving feedback to the students to make them improve their learning or use of language. So, this is the primary goal. If you will give feedback to the students, so why are we evaluating the results of the exam? So, what is it going to bring us? If you are going to depend on exams, then it is much needed.

Furthermore, to get to minimize classical written tests, we should come up with some new techniques and some less stressful ways for students to assess their learning and give them constant feedback about their improvement and involve them in the process because language is a process, we cannot test the language with one exam. It is a learning process. We do not have to test, and we cannot judge students' learning by just looking at one exam results.

Researcher: Did you find the checklist beneficial?

Mirjana: It made me aware of many things. I became more confident than before because I saw how things should be, and I judged what we were doing. It gave me some awareness about what we do and what we should do. Moreover, it gave some confidence about the good stuff we were doing unconsciously.

Researcher: What about the part you talked about your context-adopted philosophies or practices and your opinion about it? How did you benefit from such phase is this?

Mirjana: When you do these things, you are not aware of what you are doing. Once we started discussing the assessment types and how we evaluate students, what we should do, and what we should focus on, I started to see myself and how I should change certain things. It made me aware of particular suggestions that could be proposed. And I had a chance to reflect on what we did, what we do here. So, it was also a good experience for me. Now, I can think about serious proposals.

Researcher: How about the suggestion phase when you talked about your dream?

Mirjana: I think it was the best part of the research because I was free to tell you whatever I think and how it should be. Of course, it needs some improvement, and it needs some details, and it needs to be rethought. I do not know that before putting it into practice. We can never know how it would work or if it would work so well. However, the ideas I generated were satisfying for me. I would love to implement some of them or try some of them to see if they would work, but maybe not all of them altogether. It would be like quite too radical for a start. We could think of making some changes, especially starting with the writing assessment. For example, we can make it a process. It would not harm anyone, or it would not cause any significant issues. Even if you do not change it now, we can start training the teachers on how to grade papers or how to use portfolios. We can start it as a project first and then, later on, make the changes in that assessment types.

Researcher: As per your current context, could your dream be implemented?

Mirjana: Yes, it could be. I told you about my worries regarding implementing such a change. But certain things can be implemented. Yes. Why not?

Researcher: Can you negotiate it with the management?

Mirjana: I do not think so because we are not alone. That is why proposals or decision making does not only belong to only one person. I could not make such radical changes

on my own. I can only propose these kinds in our meetings, and then if they agree, we can propose something altogether.

Researcher: Maybe if you felt that there would be some benefit for the students and if you believe in it, you can change it into a report.

Mirjana: As I said, maybe not changing the exams first, we can change the project's work and see if the portfolio works for the students. We can start slowly to see the results in a small-scale project, to see how their writing improves when we go through a process like a portfolio work. If it worked and they started improving, we can propose some changes in assessments generally because in our context it is difficult to make significant changes or propose big changes. They can quickly be rejected. So instead it is better to go slowly.

Researcher: How do you think your assessment literacy could be developed more? What do you need in order to change it into something that could be practically implemented? Apart from taking part in the study?

Mirjana: I have ideas, but I can never really base it on any literature.

I need some support from literature to support my idea or my suggestion, maybe kind of professional training. For example, if I create a proposal, I need to refer to specific studies, and I can prove my points. I can argue more strongly when it comes to the proposal stage.

Researcher: when it comes to technical stuff, I mean practicality. What exactly do you need?

Mirjana; I need the support from teachers basically because the teachers and students will be involved in this process more. They should be trained. We need some training on how to design and grade the assessments and how to be objective. Students should also be somehow trained in the classroom to read through rubrics, to receive feedback, to work on feedback, to change certain things. To improve their writing, they should understand the process. Otherwise, it would be unclear to them, and they will not accept it.

In short, I need some training and some theoretical input, but I think it should start in our context. The training has to be contextualized. Everything should be the Internal as per our contextual needs.

Researcher: And when it comes to assessments, is it one of your favourites areas?

Mirjana: I have never been in a context where I was trained as an assessor. Now, when I talk to you about these things, it is interesting to me. It is intriguing. I would want to

know more. However, before that, I never thought that I would be interested in assessments or exams design. I felt the need for some training. I want to know more about how people are doing.

Researcher: Did you acquire anything from that kind of discussion during the research?

Mirjana: The checklist gave me so many ideas because it was very detailed, and it is made me think about every detail that we are going through here. Our discussions regarding the checklist were very thought-provoking. It made me think about what I was doing and how I can improve things. So, it was fruitful from my side. The other stages, you just let me come up with my ideal world. The first part you were more involved basically, but then you gave me the microphone to assist myself.

Researcher: As a teacher for more than 15 years, do you believe that such kind of contextualized research involving teachers would help the teaching and education process?

Mirjana: These discussions gave me a voice because I shared my ideas and concerns. This is how we learn. We share ideas and improve—just listening to one another even to problems without solutions because it is not necessarily that we could find a solution. Just complaining about certain things happening around us can help us find solutions because we listen and support one another. Sometimes, you feel okay if it is not just you facing this problem. This even gives you some confidence in your practices. It is like everybody is facing the same problem, so let us deal with it instead of blaming yourself for it. Teachers usually blame themselves and their lack of experience. I think these research practices mean a lot; they teach us a lot; we learn from each other because it is about reality, not theories and abstract things. So basically, this will be the most valuable exchange of ideas because we are in the same context and we have to find solutions to our com

mon problems. Solutions to our problems come from inside.

## Appendix 11: Analysis of Data Collected from The Open-Ended Questionnaire

Themes	Table 3.1 Participants' Pseudonym							Talen
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	
Theme A1.1 Theoretical Pre-service Preparation	1. BA in English literature  2. MA in Teaching Turkish as a foreign language  3. A certificate in teaching, it included a little about assessment  4. A pedagogical course on teaching skill  5. No certification or any theoretical preparation in assessment	1. BA in English linguistics  2. MA in Information science  3. No other certification or any theoretical preparation in assessment  4. " most of the textbooks I was exposed to were related to linguistics in general; nothing was about how to teach, how to test, there was nothing about that."	1. BA in English language teaching  2. MA in educational science  3. I did not have one course which focused on testing; it was a component of a methodology course but only for a few weeks; it was like a chapter or two. It was not a course strictly based on assessment	1. PhD in critical discourse analysis  2. MA in linguistics  3. BA in English literature  4. In college once I studied a course related to methodology; it was about teaching but nothing related to assessment  5. there should be a mandatory course that is given to all teachers on the college level	1. BA in Education  2. MA in TEFL  3. Phd candidate  4. No theoretical preparation or exposure to assessment. I just started exploring when I was given the task, s I had to fix the problem. Whatever I studied was focused on methodology. I haven't taken any courses during my academic preparation related to assessments	1. BA in literature  2. MA in Education  3. Currently working on my PhD in Education	1. BA in English  2. MA in Literature  3. I haven't earned any assessment certificate  4. I took a teacher training to become a teacher, they were nine modules, two of them were on assessments, but they were minimal. There was one book on how to design exams, but it was in Turkish	1. BA in English language and literature  2. No master's degree  3. I haven't earned any certificate, or theoretical preparation about assessments

Theme A1.2 Practical Pre-service Preparation	Training on teaching but nothing on assessment	"I don't remember having any kind of workshops or training on how to create an exam or any assessment before starting my teaching or assessment task."	No practical training before I started teaching	I haven't taken any workshops, seminars, training, or attended conferences related to assessment. There should be a practical training or a certificate on assessments that should be given to teachers before performing assessment tasks	No textbooks, no workshops, no seminars, no training	Before I started teaching, I had nothing to do with education because all my qualification, I started working on them many years after I started teaching because I wanted to improve myself and empower my career. I started working as a teacher because it was the easy job at that time	As part of the pre-training, we had practical workshops where we were to design very simple exams after providing us with the material.	No training. I have never done that
Theme A2.1 Theoretical In-service Preparation	Teacher development course mainly on teaching skills with very little focus on assessment.	"I don't remember taking any"	Nothing	Unfortunately, no. But I didn't have this pre-knowledge or pre-training it was just given as part of my teaching assignment. I believe that teachers should be wholly responsible for creating their assessments, whether graded or not graded because they are the ones who are teaching and know their students best.	No, at all. I was just given the task without any guidelines. It was all personal effort. There is no unit for assessment. Everything depends on you, that is the reason exams differ from one team leader to the other, it depends on each one's competency, experience, knowledge, talent, no standardization or consistency at all	In my previous context, yes during my masters' there were courses about assessment, Bloom's taxonomy and how to create questions, but in the current one no.	We had a two-day training, we enrolled, got the training, and then the certificate. It was a theoretical training, but it was not about assessments. however, I don't believe that with those two minor preparatory activities, but my experience on the job experience filtered me as an assessor	We have taken workshops but not on assessments. They were about teaching, nothing about assessment, they weren't practical ones they were like lectures

Theme A2.2 Practical In-service Preparation	<p>In a previous context, I received informal training on grading from people involved in exam design in the form of standardization meetings for standardizing grading of speaking and writing sections of the exams to ensure consistency of grading among teachers and to minimize subjectivity in grading and objective use of provided rubrics.</p> <p>In the current context, no. "I learnt by hand on assessment. I depended on instinct, experience, feeling"</p> <p>"Unfortunately I haven't received any workshops or training, but it was just personal experience shared among colleagues, and my husband is in the field, so I used to discuss all things with him because he worked in an exam unit and he knew about certain specifications."</p>	<p>"I was not exposed to any training about assessment, we were sent certain material, and we were asked to create exams. There is nothing, no guidance, we are just doing it under pressure, but how it, should be done, we are stuck with that."</p> <p>"The assumption is that if you are a teacher, you are n assessor, t is part of your job, it is taken for granted that you have to create exams, it comes naturally because their view is that exams are asking a set of questions which teachers do all of the time while explaining, revising lessons, asking questions to confirm students' understanding."</p> <p>"The only practice that we do after we create assessments is that we send it to the head of the department, who is the reviewer to check and edit our work. We learn a little bit from the feedback taking into consideration that whoever is giving the feedback is not trained</p>	<p>Yes, but not in my current job. Because I started working in an exam unit in one university, I was lucky enough to be trained to assess. It wasn't a theoretical one. It was a practical one, we were sent to a few workshops, seminars related to assessment, but they were general ones like guidelines. Then in this testing centre, they started to train me how to base assessment on curriculum, taught material, and questions are geared towards them. Again, we put the demographics of students into consideration. I was also in the grading team and testing team but in my current context, I received nothing. You are just required to perform this task, and it should be done with no input, no care, and I feel that something should be done."</p>	<p>They did not give much attention to training on assessment; most of the training we received was on methodology. It is taken for granted that if you take training on teaching methodology, you can assess. This is a lack of knowledge or professionalism by the institution. There should be training on assessments before involving teachers in such tasks to be familiarized with the purpose of adopted assessment practices, types of assessments. The least attention is given to that skill despite being the most crucial step in the whole education cycle plus it does not only assess students, but it is an assessment for teachers' and institution performance as well. Assessment is where you end up; it is the end of the journey; it can affect institution reputation. It is reflecting the effort of all involved stakeholders. In my context assessment id ignored and looked over. There is a lack of knowledge among assessors. The problem</p>	<p>No, unfortunately, I asked for it, but I did not get it. I needed practical practice and guidance. I did not want theoretical ones but no response. I depended on myself. I read but that much because we don't have time. I needed someone with experience so that they can give me a hand. I needed someone to teach me how. The problem that my supervisor always approves my exams and I have never received one single comment, and I know that I need development and assistance. The problem that they are judging based on their standards; they don't see that I need it. I believe there should assessment unit involving teachers with less load because assessments take time, but they should be teaching to be aware of the students and their needs, levels, and preferences and I don't think all teachers should be</p>	<p>In my previous job, I attended several workshops. In the current one nothing; It is just personal effort.</p>	<p>In the current context, I haven't taken any practical assessment preparation before performing my assessment task. I depended on my previous experience in other places. I tried to work on myself. I read books on assessment and testing. I became a member of the testing and evaluation committees. I read about the common European framework. I was interested in testing vocabulary, and I compared between testing the vocabulary traditionally and using their framework. I checked online courses; I came across one about evaluation. I trained myself on IELTS exams, I took the exam myself, I got books, I trained myself so that I can use it to train my students to pass it, but I was never involved in designing standardized tests, just as a user and a trainer.</p> <p>I think we need to know more about their design because we can learn a lot about styles of checking reading and listening; it was quite guiding for me.</p>	<p>I just learnt from other teachers how to do it. I learnt by apprenticeship.</p> <p>They just give us things we have to include in the exams, like what o keep in mind but training or any practical practice on how to create assessments, no not at all. I guess they don't care about this despite its importance in evaluating all educational elements because everything is ready now, if you are using a specific textbook, they provide you with its assessments and they market their philosophies and ideas. Everything is ready no need to create it yourself. I somehow agree. I can be involved in creating maybe simple quizzes to check their understanding, but for summative</p>
--	--	---	--	---	---	--	---	---

	<p>" I also tried to educate myself. I searched online for rubrics because my current context provided nothing, I just tried myself because I wanted to perform well, I did it intuitively based on my experience as a teacher and a student who sat for various assessments."</p> <p>" We put the parameters; there were not any guidelines, they just ask us to do it, we came up with our own, we just followed a sample given to us, we were not given any guidelines, training on how to create assessments."</p> <p>For example, I learnt from teaching courses about the communicative approach, so when I create exams, I believe I should not test them in grammar rules with multiple choices because it won't help. No one taught me that I just used my teaching experience as an</p>	<p>as well for such a task."</p> <p>" The feedback we get focuses on avoiding being repetitive or adding too easy or too challenging questions, making sure the questions are related to the curriculum, so you can say we learn about assessments by hands-on assessment, we develop our assessment skills based on the little feedback we receive and our reflection on our work, our intuition, and our previous experience with assessments as students, also our experience and contact with students, knowing the culture, knowing our students' motivation, attitude, and culture, but unfortunately not based on something solid"</p> <p>" We have an exam unit, but it is not concerned with the content of the exam, tools, exam creation, matching learning outcomes to assessments, it performs an</p>		<p>that we don't have the power or a say. You just do things as being asked to do.</p> <p>Unfortunately, on a personal level, I didn't do any steps to develop myself in terms of assessment. I believe teachers are scared of assessment. It is a little intimidating for them because they are scared of not satisfying students, so they did not want to take responsibilities or be part of that they don't want to get the blame, so they prefer to leave it for others. I prefer formative assessments. I also avoid the summative assessment. It is a huge responsibility unless you are qualified to do it and have the knowledge, talent and experience to fire back and defend your point. We avoid it because it comes with its problems. If you are not professional, you can harm students, especially that we are doing it among other piles of tasks. To do it, you have to be devoted. We are overloaded with</p>	<p>involved some teachers are not competent at all and not interested in the task. Those who have the knowledge base, motivation, interest, experience, need to be trained, guided, given quality preparation not by anyone but by professionals and expertise.</p>		<p>Traditionally, best exams were the difficult ones which make students suffer. I believe that my job as an assessor is not to make students suffer or feel miserable, but to feel the achievement. I got this by experience; we should focus on learners' skills and capabilities. We should be shining them not burning them. But how I need to have this knowledge, I need someone to give it to me. I need the knowledge to be able to critique what is around me professionally. I need to know about both classroom and standardized ones. When students perform less than what we predict, we always blame them; maybe it is us or our assessments. We need this literacy to judge and critiques the whole situation professionally and fairly.</p>	<p>tests, I believe it should be provided by them simply because we are not prepared for the task.</p>
--	---	--	--	---	---	--	---	--



	input for my assessment experience.	administrative role in the form of receiving exams, storing them, distributing them to invigilators and back to graders without any guidance."		tasks; how we can be fair.				
--	-------------------------------------	--	--	----------------------------	--	--	--	--

Theme B1 Precise Role	I started as a writing team leader responsible for creating and designing writing assessments. Then I was promoted as the head of the foundation program and responsible for revising, auditing, approving, and in many cases designing of assessment in the foundation program. With other team leaders or course moderators, we are responsible for designing summative exams (mid-term & final), quizzes, assignments, project deliverables, and presentations tasks, in addition to designing answer keys and rubrics for grading all varied types of assessments."	I am a team leader for listening. My role is to create exams, quizzes, assignments, projects, presentation mid-term and final exams related to the module I am assigned as a team leader for; the module could be reading & vocabulary, writing, listening, speaking, grammar, 'We have design assessments, select audios for listening and texts for reading that are appropriate for the learners' levels and come up with questions or tasks similar to the ones we teach in the textbooks and practise with students before. We also look at the learning outcomes that we are targeting and find questions that can help in assessing them.	I prepare assessments for the two levels of the English preparatory program. I prepare reading texts with multiple-choice, true/false, insertion questions. I prepare the writing component of the test which may focus on sentence construction, paragraph construction, and essay construction depending on the level of the students They all required to combine sentences, use transitions and conjunction. Apart from the summative assessments, I create projects with deliverables, activities, and rubrics. I am involved in both summative and formative assessments.	I am involved in creating presentations, projects, mock quizzes, in-class assignments but not summative ones. I create grammar, vocabulary, reading and writing quizzes with multiple-choice questions. For speaking, I created a checklist for assessing students' presentations.	Team leader and assessor. I create formative assessments like quizzes, assignments, projects, deliverables, presentations and summative ones like mid-term and final exams. While designing, I pay attention to the content of the course, type of questions or tasks I need to include, learners' levels and many other things. It takes time; it is not just putting things together. Usually, when you are rushed to do things and finish them within a minimal time, it affects the quality of whatever you deliver.	I am a team leader, and part of my role is to create assessments in addition to rubrics and answer keys for grading. I create assignments/quizzes for the four skills in addition to grammar and vocabulary in addition to projects with their deliverables description, presentation, mid-term and final exams.	I am a team leader; I create quizzes, in-class assignments, experiential learning activities, projects, presentations and mid-term and final exams.	I create in-class assessments, quizzes, assignments, projects, presentations, but I am not involved in summative ones. I depended on the internet, previous exams, the textbook, my knowledge and experience.
--------------------------	---	--	---	--	--	--	---	---

Theme B2 Selection Criteria	<p>I was the team leader for writing, and whoever is given this assignment should be responsible for all related assignment "it comes with the assigned role'. My second role as a reviewer and auditor of assessments also came with the new role. "It is part of the job description of the head of the department to review and approve all forms of assessment designed by team leaders or course moderators'</p> <p>"The selection is not based on certain criteria related to assessment. It was a role-based."</p>	<p>I started working as an instructor, the head thought I had leadership skills, she assigned me as a team leader, and the assessment tasks were part of my role as a team leader for a module. It comes as a package; whoever is a team leader for a module should be creating assessments for that module. There are other tasks for a team leader lie leading the instructors teaching the module, conducting weekly meetings with them, supervising new instructors and mentoring them, observing and evaluating teachers, choosing material and textbooks, distributing syllabus and coming up with supporting materials, deciding on course objectives and learning outcomes and finally deciding on assessments to assess the learning outcomes.</p>	<p>They just approached me because they wanted to change the assessment team, and they asked me if I was willing, and I accepted, but it was not based on specific criteria.</p>	<p>It is a kind of an assigned task by the head of the department.</p>	<p>I am the team leader, and it's my responsibility to design assessments. We don't have an exam unit or specific criteria for choice. I create assessments, and the senior reviews them.</p>	<p>Based on availability, because there are no criteria for selection. it is done maybe randomly</p>	<p>I was just assigned the role because I taught this before, I had some suggestions that I shared before maybe this is the reason. Maybe because of the long experience. But there are no specific criteria for selection of instructors to be assessors.</p>	<p>Based on my experience as a teacher. I was chosen randomly based on likes and dislikes.</p>
Theme B3 Years of involvement	<p>I started teaching 17 years ago, but I was involved in the current assessment role for seven years</p>	<p>As an assessor, even years</p>	<p>Eight years</p>	<p>15 years</p>	<p>Nine years</p>	<p>Since I started working in this contest before I was not involved, around six years</p>	<p>15 years</p>	<p>12 years</p>

## Appendix 12: Analysis of Data Collected about the Participants' Confidence Level with Assessment Knowledge Base

### Analysis of Data Obtained from the Open-Ended Checklist (Competence Section)

	<b>Table 3.2</b>							
Sub Themes	1. Participants' Responses Regarding their Level of Confidence with Disciplinary and Pedagogical Knowledge							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Knowledge of Second Language Learning Theories:	<b>Medium</b> I was exposed to these theories after graduation during my pedagogical training. I built this knowledge based on my own experience as a teacher and how students learn.	<b>Medium</b> I am not sure I learnt about all of them. I learnt about language acquisition.	<b>Highly</b> Undergraduate and MA studies	<b>Medium</b> I was not exposed to them, but I have some knowledge about it.	<b>Highly</b> I learnt about them during BA and MA.  I refer to them in my teaching experience with problematic cases who have difficulty in learning.	<b>Medium</b> Not so much because I haven't studied about them before, I just read about them.	<b>Highly</b> From my theoretical and practical preparation + experience and hand-on learning	<b>Medium</b> I had heard about some of them, especially when I started teaching.
Knowledge of Second Language Teaching Theories	<b>Highly</b> I worked on it during my pedagogical learning. I was taught how to teach, teaching styles. I applied this in my teaching career because, in our educational system, these things are given more importance.	<b>Highly</b> I am quite confident not only on the theoretical level but also as a result of teaching I have been teaching for a long time now. Also, I was exposed to training about methodology through trainers in my working place.	<b>Highly</b> Undergraduate and MA studies	<b>Low</b> No theoretical preparation, no training, no workshops about them. I approach students in class by intuition, experience, my experience as a student, what I feel is the best way for them.	<b>Highly</b> I have been exposed to all of them during my master's degree because mine was in teaching English as a second language. I had to use during the practical part and pre-service training.	<b>Highly</b> Very confident. I got it in the master and from cumulative experience from reaching, reading and research I have done.	<b>Highly</b> For 25 years and I have been trained with these methodologies and theories. Theoretically and practically, I have used them.	<b>Highly</b> Highly from workshops I attended
Knowledge of Learning Strategies	<b>Highly</b> We were taught about learning strategies and how to use them to make learners learn a language plus from my experience as a teacher also.	<b>Medium</b> I use them automatically while teaching maybe theoretically I am not aware of them, but when I look to the examples you are adding I can see I apply them by experience but not as a professional	<b>Highly</b> Undergraduate and Ma studies. Prepared theoretically and by experience	<b>Medium</b> I am familiar with them. I do them unconsciously based on my knowledge as a student and how I was being taught, what goes well and what does not go	<b>Highly</b> Taught as part of my master's	<b>Highly</b> I used them all intrinsically without even learning about them when I started teaching, and I started developing my knowledge about them, and again when I started my master's, so I got to apply them, and they became a kind of need. I think even if teachers didn't study theories or	<b>Highly</b> Experience + preparation	<b>Highly</b> I use all of them by experience, and I learnt them a long time ago.

				well. I use most of them.		methods of teaching, they would get it by experience.		
Knowledge of Learning Styles:	<b>Highly</b>  I was taught, and I got them by experience. I had some workshops about learning styles in my working context. Also, I depended on my knowledge about how to learn.	<b>Highly</b>  I am aware of this theoretical and experience.	<b>Highly</b>  Proposed theoretically and prepared for them	<b>Low</b>  I was not exposed to them. I do this unconsciously. I try to combine them all to address different styles in class in terms of teaching, but in terms of assessment, I am not confident in addressing them in designing exams.	<b>Highly</b>  I know them, and I have even taught them in one course I taught before	<b>Highly</b>  I apply them all. During the same class, you appeal to different styles by using different tasks. Of course, by experience	<b>Highly</b>  Experience and theoretical preparation	<b>Highly</b>  From experience but learning no
Knowledge of Theories of Language Use (Language Models)	<b>Medium</b>  I don't know the names of the theories, but I know their components.  I was not exposed to this knowledge, but I know by experience on which component I should focus when we are teaching or testing a language.	<b>Highly</b>  competent through theoretical preparation and experience	<b>Highly</b>  But I have my own beliefs about them. As a native, I had no idea about grammar until I become a teacher. When I started teaching, I started to learn grammar in order to teach this to my students.	<b>Highly</b>  I would say confidently because this is my major, education-wise. I am a linguist.	<b>Highly</b>  Master's degree	<b>Medium</b>  In the classroom, we focus on the knowledge of grammar, vocabulary, sentence structure. We don't focus on the linguistic aspects, discourse analysis, communicative or strategic competences or higher linguistic levels because of different contextual factors or philosophies that controls our teaching. I took them as a student.	<b>Medium</b>  By experience I am, but by the theory, I am not that confident. I have my ideas filtered through language teaching, what is language teaching, and what should I teach.	<b>Medium</b>  Not all of them. I learnt some of them in university.

Sub Themes	2. A. Participants' Responses Regarding their Level of Confidence with Knowledge of Assessment Purpose, Content, and Methods							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Knowledge of language testing and assessments history	<b>Low</b> I just know how I was tested and how things changed now.	<b>Low</b> I was never exposed to this.	<b>Medium</b> Not that much	<b>Low</b> I was not exposed to that.	<b>Low</b> I didn't get any assessment preparation. I wasn't taught about it or exposed to it within a formal course.	<b>Low</b> We hit on it during the masters'	<b>Medium</b>	<b>Low</b> I haven't studied it
Knowledge of language testing and assessment philosophies	<b>Low</b> I was not exposed to the literature about it	<b>Low</b> I was not exposed to this before. All I know is that I assess because we want to make sure that learning took place and learning outcomes were achieved, that the student reached a certain level that we can allow them to move to the following level. This is the main purpose of assessment in my context.	<b>Highly</b> Theoretical and practical preparation during BA and Master which were in education	<b>Low</b> Lack of exposure	<b>Low</b> I don't know about them.	<b>Low</b>	<b>Low</b> I am not that confident I would just depend on my experience.	<b>Low</b> My teaching experience
Knowledge of the Impact of adopting testing or Assessment Philosophy	<b>Low</b> Again, not exposed to it but I can guess them by experience.	<b>Low</b> I was not exposed to this.	<b>Medium</b>	<b>Low</b> I know generally, but I don't know exact and deep impacts. I can feel things. I know them by experience but not professionally. I am in the middle of everything. I can feel the impacts of adopting certain assessments approaches.	<b>Medium</b> Based on experience. I have seen the impact practically as a teacher and assessor. For example, I have seen when a country adopts a certain specific one method of assessment, and they are not varying their methods or addressing different needs of students, and the end target is the scores, and therefore I can see the impact on the assessment.	<b>Medium</b> Medium. I do understand what you say I can feel how it affects the level of students and in the long run, what would be the ramification of these philosophies.	<b>Medium</b> Again, by experience	<b>Low</b> All I know is that it will control the tools I am using.
Knowledge of Major Assessment Qualities: Authenticity Interactiveness Practicality Washback Fairness Ethicality	<b>Medium</b> Based on my experience, but I was not taught. I just figured it based on my experience or our discussion as an assessor team.	<b>Medium</b> I think I know them in practice, but I was never exposed to them, in theory, I think we are somehow applying them.	<b>Highly</b> Theoretical and practical preparation during BA and Master which were in education	<b>Medium</b> I was not exposed. I know them by experience, practice, involvement with the students.	<b>Medium</b> I try to use these qualities, and I pay attention to all of them when I am designing an assessment.	<b>Medium</b> Medium because in my context, they believe it does not matter if assessments show the application of skills as much as they show grades to help learners pass to start their majors.	<b>Medium</b> I was not trained or supported, theoretically. I feel them I do them intuitively. No formal education	<b>Medium</b> By Experience

Knowledge of Assessment Types (Traditional/Alternative) (Norm- /Criterion-Referenced) (Summative/Formative) (Externally /Internally Mandated) (Standardized/Classroom)	<b>Medium</b> I know and do some of them, but I am not familiar with the literature or terminology. I am confident, based on my experience.	<b>Low</b> No, I don't know about them.	<b>Highly</b> Theoretical and practical preparation during BA and Master which were in education	<b>Low</b> I am not familiar with them. I know the meaning, but I don't know what actually they involve or how to design or apply them professionally.	<b>Medium</b> I experience some of them, but I did not know the definition of some of them.	<b>Medium</b> I totally know the difference. The type I am involved in is not standardized. I use a mixture of summative and formative ones.	<b>Medium</b> I can figure out what is going on through my experience.	<b>Medium</b> I knew now after you explained. I noticed that I do them by experience. I haven't learnt them.
Knowledge of Assessment Methods:  (Tests, Portfolios, Performance Assessment, Self-assessment, Peer assessment, Checklists, Interviews, Questionnaires, Role-plays)	<b>Medium</b> Based on experience	<b>Medium</b> I do them by experience hand on assessments.	<b>Highly</b> Theoretical and practical preparation during BA and Master which were in education	<b>Medium</b> By practice, I use them but wasn't prepared to design them.	<b>Medium</b> From my experience as a student. I was exposed to many of them. I was not taught, but I understood how to deal with them because we were given the rubric of how these assessments would be graded and what was expected from us as assessment takers.	<b>Medium</b>	<b>Medium</b> I know what they mean, I can use them, but I would love to have someone tell me how to do them professionally.	<b>Medium</b> Familiar per experience
Knowledge of Major steps in Language Test Development : (Test Purpose, Construct Definition, Content Specification, test Specification, Item writing)	<b>Medium</b> I was not taught these things. I cannot name them, but I do them while designing. I connect the items to the learning objectives, but I think it is not enough. I am doing them intuitively. It should be more advanced. " We follow the model."	<b>Medium</b> We go through those steps but not formally, or technically I know how to detect the purpose and the content from the syllabus, but I haven't used test specification. They are done intuitively. I was never exposed to such preparation, item writing, for example, which one to use the pitfall of each one when it is best used.	<b>Medium</b> May be construct definition, I would need more about it.	<b>Low</b> No, at all. I do them intuitively. I am not even familiar with some of the terms like construct definitions and specifications.	<b>Medium</b> Again from my experience as a student; this was like an extra course in my master not mandatory, and I volunteered to take it because as I teacher I knew that at some point I have to make an exam.	<b>Medium</b> The steps maybe. The terms I am not sure of. I use them intuitively when I design the exams, but I was not trained on them. They came by experience. I developed my own expertise in assessment design based on my teaching experience, knowledge of the student and the context. I hope it is going the right way.	<b>Medium</b> I did most of them intuitively and by experience. I was not prepared. I can set the purpose and link it to course objective, but test specification not really competent in that.	<b>Low</b> I did not know about that. I understood when you explained.
Knowledge of Cognition Taxonomies Required for Different test items or Tasks  (Bloom's / Webb's)	<b>Medium</b> I know them by experience. I cannot say I did this because I followed this taxonomy. We are not that professional. I cannot say I am a tester or an assessor because I was not trained, taught, supported. I was just given a task, and I dealt with it myself. I feel it, I do it but not in a standard/formal/ manner. What I do I feel is	<b>Medium</b> I do it intuitively, but I have not been exposed to taxonomies. I know that there are straight forward questions, and there are questions that test critical thinking. I am not aware of all levels. I design questions where students need to conclude, which is a higher thinking level, not everything is	<b>Highly</b>	<b>Low</b> Zero-knowledge about that	<b>Highly</b> I know them as I taught them as one component of a course I taught. When I create my exams, I design my items in a way that addresses the different levels.	<b>Highly</b> I took it a part of my master's	<b>Highly</b> I am good at this. It was self-study. I got it in my teaching courses as well.	<b>Medium</b> By experience

	right, but it could be much better if we had some kind of training or if we could relate things to one another.	on the memory or understanding level.						
Identify and state language assessment purpose.	<b>Highly</b>  Experience  It is already set by the context.	<b>Highly</b>  By practice plus these are already set for us.	<b>Highly</b>  By education and practice and experience	<b>Highly</b>  By experience and practice	<b>Highly</b>  It is from experience.  I know the purpose. It is already set for us.	<b>Highly</b>  By experience. The assessment purpose are already set for us.	<b>Highly</b>  Yes, I can out of my experience.	<b>Medium</b>
Write the goals and objectives of instructions and consequently, their assessments.	<b>Highly</b>  We can do it as a group in the form of discussion as an assessor team. We can depend on our personal judgment. Goals of instructions and assessment are set in the syllabus and textbook imposed on us by a higher committee.	<b>Highly</b>  By Practice. Learning outcomes are set for us.	<b>Highly</b>  By education and practice and experience	<b>Medium</b>  Not much. It is an intimidating task for me. It is difficult for me. I need to be guided.	<b>Highly</b>  I got it from my teaching experience because for every course. The objectives are set, and we have to cope with these goals or objective and adhere with them.	<b>Highly</b>  By experience. My teaching is guiding me. Not all of the time the objectives are set in the textbooks we are teaching, sometimes I do them myself.	<b>Highly</b>  It is again out experience and theoretical knowledge.	<b>Medium</b>
Align curriculum objectives, instruction, and assessment	<b>Highly</b>  By experience. As a team, we follow a plan prepared by a higher committee, and we have samples.	<b>Highly</b>  Every assessment we create, we have to match it with the learning outcomes provided by the context.	<b>Highly</b>  By education and practice and experience	<b>Medium</b>  Sometimes I can do it. Again intuitively	<b>Highly</b>  Teaching experience	<b>Highly</b>  By experience. It is there in the test books, and I just need to align them to my assessments.	<b>Highly</b>  From experience	<b>Medium</b>
Define the language construct(s) an assessment will give information about	<b>Highly</b>  Through experience	<b>Highly</b>	<b>Highly</b>  By education and practice and experience	<b>Highly</b>  From a teaching experience, I would say yes. The construct is what I am testing, and it is derived from the learning objectives.	<b>Highly</b>  I learnt them by practice, but I was not taught.	<b>Highly</b>  By experience	<b>Highly</b>  From experience	<b>Medium</b>
Utilize alternative means of assessments to make decisions based on substantive information	<b>Highly</b>  It is adopted in my teaching context, and I also believe in it. I initiated some more alternative ones based on my initiatives.	<b>Highly</b>  By Experience	<b>Highly</b>  By education and practice and experience	<b>Highly</b>  By practice	<b>Highly</b>  By experience and practice	<b>Highly</b>  By experience	<b>Highly</b>  I do it out of my experience.	<b>Highly</b>
Construct test specifications/blueprints to design parallel forms of a test	<b>Medium</b>  I used a simple one for reading. I did not initiate it. It came with the package. I changed certain things because it came from another department.	<b>Low</b>  I don't know	<b>Medium</b>  By education, practice, and experience. I don't use them in my current context.	<b>Low</b>  No at all	<b>Low</b>  I never used them. I know the terms, but I was never trained to use them.	<b>Medium</b>  We use a form in my context, but I guess this does not reflect what you explained by the test specification.	<b>Low</b>  No, I don't know that. I was not exposed to it.	<b>Low</b>



Write selected-response items such as multiple-choice-, true-false, and matching.	<b>Highly</b> By experience	<b>Medium</b> By experience	<b>Highly</b> By education and practice and experience	<b>Medium</b> I am doing them intuitively but not professionally. I don't have the knowledge or input.	<b>Highly</b> Again, by practice and experience. Working once with a testing centre helped me because I was training candidates to deal with them, so it helped me when I started designing my own tests.	<b>Highly</b> By experience	<b>Highly</b> I do them, but again I need the techniques, I got some techniques during my teaching training, but still, we need more.	<b>Medium</b> By experience
Design constructed-response items (for speaking and writing)	<b>Highly</b> By experience	<b>Medium</b>	<b>Highly</b>	<b>Medium</b> By Experience. I need to know techniques of how to be objective and fair.	<b>Medium</b> With writing, yes but for speaking no I am not familiar with it	<b>Highly</b> By experience	<b>Medium</b> By experience	<b>Medium</b> By Experience
Construct well laid-out and perfectly legible items/tasks	<b>Highly</b> By experience	<b>Highly</b> I try as much as possible to make them as clear and possible.	<b>Highly</b>	<b>Medium</b> By experience	<b>Highly</b> I learnt about face validity, From the feedback of my supervisor. I compare mine to other exams. when I am invigilating other exams, I check how students' respond and what is difficult or user friendly to them	<b>Highly</b> By experience	<b>Highly</b> By experience	<b>Highly</b>
Provide clear, explicit, unambiguous instructions	<b>Highly</b> By experience	<b>Highly</b>	<b>Highly</b>	<b>Highly</b> I do it by experience. I try to be specific.	<b>Highly</b> From experience and students' feedback and their questions makes me learn and develop. This is so much related to your teaching experience and dealing with students.	<b>Highly</b> By experience	<b>Highly</b> By experience	<b>Highly</b> By experience
Provide examples to make candidates familiar with assessment techniques	<b>Highly</b> By experience	<b>Medium</b> No, I don't do this, but I can do it. I believe if we train them enough and they have similar examples in the book or the material I don't need to do so	<b>Highly</b>	<b>Highly</b> It depends on the difficulty of the task.	<b>Highly</b>	<b>Highly</b> I know how to do it, but I don't do it.	<b>Highly</b> I do it	<b>Highly</b> By practice
Include many items to enhance scoring reliability	<b>Highly</b> By experience. We use many items in the tests, and we test in different ways.	<b>Low</b> I am not sure if I am for this or against. We had an argument about how long the assessment should be when it comes to items, and I don't know.	<b>Highly</b>	<b>Medium</b> By experience, by practice. I need to know how to do it without being redundant. I don't have theatrical knowledge.	<b>Highly</b> From practice and experience, we have done so many exams as needed by the organization.	<b>Highly</b> By experience	<b>Highly</b> Yes I know that	<b>Highly</b> By Experience

Write test syllabuses to inform test users of test formats where applicable.	<b>Medium</b> We don't use anything formal. We prepare students in the class for test formats so students can get ready for the type of questions.	<b>Medium</b> They are aware of the task	<b>Medium</b> By education and practice and experience	<b>Low</b> No	<b>Low</b> We should have a system that requires this to be done. We can't work by our self.	<b>Medium</b> We do this in revision classes before exams.	<b>Medium</b> I haven't done it myself, but I use it.	<b>Medium</b>
Design assessments that are valid not only in terms of course content but also in course tasks	<b>Medium</b> Through experience	<b>Medium</b> I don't think this stuff is checked in my context, which cause some kind of confusion in certain tasks. We don't have someone experienced enough for setting the construct and matching the items with the construct. I feel this is how it should be done, but I could not explain why it is valid or if I am right or wrong.	<b>Medium</b> By education and practice and experience	<b>Medium</b> I do it, but I need the knowledge to validate what I am doing.	<b>Medium</b> This is from practice and experience, and I read some article, but I need to ensure my practices.	<b>Medium</b> I sometimes test the same thing in different ways. this variety is done to ensure validity	<b>Medium</b> Yes. I am doing something in class, and I test to see the impact of what I am doing.	<b>Medium</b> By practice
Design assessments that are reliable, authentic, fair, ethical, practical and interactive	<b>Medium</b> By experience	<b>Medium</b> I am trying. I am not sure when I design a test if it is reliable or valid because I am doing it based on my experience and knowledge.	<b>Medium</b> If given the opportunity, yes.	<b>Medium</b> I can't guarantee it. I am doing it by experience.	<b>Medium</b> I got it by experience I am designing, and I am reviewing my own exams, and sometimes I compare them to others.	<b>Medium</b> Based on my experience I believe I do	<b>Medium</b> I know them, but I don't know how to apply them. I need practice. Out of experience, I can say I design reliable exams, but I cannot base that to well-defined criteria.	<b>Medium</b>
Incorporate technologies in assessing students	<b>Highly</b> We use some platforms for students to submit their deliverables. We were in a stage where students did their writing exams, submitted drafts, and received feedback on the computer, but then the institution stopped that.	<b>Highly</b> I am trying to update myself. I ask them to create videos as deliverables. I asked them to record what happens behind the scene. I use Turnitin, Moodle, and PowerPoint.	<b>Highly</b>	<b>Low</b>	<b>Highly</b> It started again from the testing centre I worked at. I learnt how to deal with different technologies, using labs, and different software. Also, in class, when we are training them to give presentations, we teach them how to incorporate technology.	<b>Highly</b> Online tests and others. yes sometimes I do	<b>Medium</b> I am doing my best may be due to age we were not trained in technology. I am used to pen and paper, But I am using some. I am training myself. I use Moodle, Turnitin, whatever is required.	<b>Medium</b>

Sub Themes	3. A. Participants' Responses Regarding their Level of Confidence with Knowledge of Assessment Grading							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Knowledge of grading rational (norm-, criterion-referenced)	<b>Medium</b>  Experience	<b>Medium</b>  Experience and working with other assessors	<b>Highly</b>  My education and theoretical preparation. I got this over time when I was first preparing tests in a testing centre.	<b>Low</b>  I don't know how they are done.	<b>Low</b>  I haven't been exposed to it, but when it is done, it should be done by people who really know by professional, not people who read about and come to teach me. I need people who already applied it, know the pitfalls of things, how to do it, the best ways, what to void. I need someone who knows.	<b>Low</b>	<b>Low</b>  I don't know about it. No exposure	<b>Low</b>  I just know that they are different. I don't know about them.
Knowledge of methods that change scores into meaningful composites (points, per cent, grades, proficiency levels).	<b>Medium</b>  Experience	<b>Medium</b>  Experience and working with other assessors and maybe checklist is given by the institution.	<b>Highly</b>  Theoretical and practical preparation	<b>Low</b>  I don't know how they are done.	<b>Medium</b>  Maybe the first two I know about; I was not exposed to the others.	<b>Medium</b>  I know this by practice	<b>Medium</b>  May be not fully fine, but I know somehow out of my experience.	<b>Low</b>  Not very competent
Knowledge of basic linear scaling, weighting scaling, weighting components, and precision of results).	<b>Medium</b>  Experience	<b>Medium</b>  Experience and intuition	<b>Highly</b>  Theoretical and practical preparation	<b>Low</b>  No. I am not familiar with the terms, knowledge, or practicality, how to do it. I do it but not based on specific criteria.	<b>Low</b>  Not exposed	<b>Low</b>  I don't know about that. What I do is that I don't allocate high grades for items because this might be helpful for students.	<b>Low</b>  No at all	<b>Low</b>  If I know about this, I could fight for it because sometimes I feel something is wrong by I can't tell why. I need to know more about this.
Knowledge of scoring techniques for objectively-marked testing	<b>Medium</b>  Experience	<b>Medium</b>  Experience and intuition	<b>Highly</b>  Theoretical and practical preparation	<b>Low</b>  Again, I do it intuitively not based on techniques.	<b>Low</b>  Not exposed	<b>Medium</b>  I feel I need more. I would say this is	<b>Medium</b>  Based on my experience and intuition. When you grade, you have an objective in mind, a rational in mind	<b>Low</b>  No, unfortunately, but it is very important. If I know, I can argue about certain things.
Knowledge of the nature, purpose, and design of scoring rubrics for subjectively marked tests	<b>Medium</b>  I design them, but I believe they should be given more importance. We should have more standardization meetings.	<b>Medium</b>  Experience and intuition plus the institution guide and requirements	<b>Highly</b>  Theoretical and practical preparation	<b>Low</b>  Again, I do it intuitively not based on techniques.	<b>Medium</b>  I design rubrics again by practice, experience, feedback. I have seen people doing it, and the problems and issues and I have tried to work on them and develop them. Self-learning	<b>Medium</b>	<b>Medium</b>  Out of experience	<b>Low</b>  I just follow my feelings and intuitions.
Knowledge of consistency and moderation principles for judgment-based assessments.	<b>Low</b>  Basic knowledge	<b>Low</b>  I am not sure	<b>Highly</b>	<b>Low</b>  I do all of this by practice. In terms of knowledge, I was	<b>Low</b>  I have never learnt about it	<b>Low</b>	<b>Low</b>  I am not competent at that because I never did it.	<b>Low</b>  I don't know the techniques.

			Theoretical and practical preparation	not exposed to anything related to grading.	No source we cannot find it that easily.	I don't do that, and I don't know about it. I am not involved in this.		
Construct scoring schemes that quantify learners' performance into useful information for decisions about learners, classrooms, schools and districts	Medium  I do it based on my experience.	Medium  Yes, I do it based on my experience, instruction, checklist, guidance.	Medium  This is basically categorizing	Low  Zero-knowledge	Low  I am not doing this. I was never exposed to this.	Medium  By experience.	Medium  Yes, I do in practice, but they are based on academic, theoretical, real valid concepts? I can't tell or stand for my point, and I don't believe that what I do is totally scientific.	Low
Design scoring keys with explicit instructions for closed-ended responses (right/wrong, checklist, multipoint methods including rubrics and rating scales).	Highly  We do that by experience	Highly  Yes I do it based on experience, instruction, checklist, guidance Yes I do	Highly	Low  I do it, but I am not sure if what I do I right or wrong.	Highly  By experience, I did it a lot.	Highly  By experience.	Highly  I do it by practice	Highly  I can by experience
Design scoring rubrics with explicit instructions for open-ended responses (holistic, analytic, primary trait scoring)	Highly  We do the general ones.	Highly  Yes, I do. Through practice	Highly  We are using them a holistic way. Before, they were more analytical ones and detailed.	Low  I am not even familiar with these terms.	Highly  I am using all of this. I got it by experience and self-learning, but I was never taught.	Highly  By experience.	Highly  I do it again by practice and through experience.	Highly  I can by experience
Calculate consistency of scores/ standard error of measurement to assess scoring reliability for both closed-ended and open-ended responses	Low  We started doing that.	Low  No idea	Low  I would need to read up about that again.	Low	Low  Not at all. Never done it	Low  No, that is very specific.	Low  Low confident. I do some of this for my self-satisfaction. No one asks for them.	Low  No at all
Calculate scorer reliability through scorer reliability co-efficient to quantify the level of agreement among scorers	Low	Low  We don't do that	Low  I need to know more	Low	Low  Never used them	Low	Low	Low  No
Design training workshops for scorers on acceptable approved responses	Low	Low  I myself need a workshop on that.	Low  For myself, yes, but to design for others, I don't have that much confidence.	Low  No. I need a workshop myself.	Low  Never	Low  We don't do that as a formal practice. It is all random discussion.	Low  I haven't done it	Low  I need workshops on myself.

Sub Themes	4. Participants' Responses Regarding their Level of Confidence with the knowledge of Providing Feedback							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Knowledge of purposes and principles of feedback	Highly	Highly	Highly	Medium	Highly	Highly From my teaching experience, plus I believe in the power of feedback.	Highly I do highly based on my teaching preparation and experience.	Highly Experience
Knowledge of different types of feedback (descriptive, evaluative, supportive) with their respective functions	Medium	Medium I am not sure I know them; in theory, I know what they mean.	Highly	Low I don't have the theoretical knowledge.	Low The names I am not familiar with	Medium Yes, but we don't have the time to customize feedback, so we just give holistic general to everyone. We don't personalize	Medium I knew this through my teaching experience. Intuitively I know	Low I don't know how to use them.
Knowledge of the strengths and weaknesses of various kinds of feedback (task, process, metacognitive, and self-oriented) in facilitating student learning	Medium	Low I am not sure	Medium	Medium I don't know how to apply all those techniques.	Low Never exposed to this	Medium From my teaching experience. I know about the theoretical background.	Medium	Low Not very well
Setting educational goals and constructing tasks that will move learners towards these goals.	Highly Experience	Highly It is all about learning outcomes which are set clear by our context.	Highly	Medium I can't set goals, but I would create tasks to already set goals.	Highly I use it from my teaching background as a teacher and supported by my training as a teacher.	Highly From my teaching experience. I know the concepts.	Highly Out of teaching experience	Highly Experience
Communicating to learners the criteria for success through different communication modes like telling, showing, and having learners discover	Highly Experience	Highly We usually do that at the beginning of the semester, and even before each assessment, we tell them.	Highly	Medium I feel the importance, I know the concept, but I don't know how to apply them.	Highly I do this	Highly From my teaching experience. I do it intuitively.	Highly Competent from teaching experience	Highly I can do it
Showing learners, the gap between their current stage of development and the next/final stage of development	Highly Experience	Highly Sometimes I compare their competence at the beginning and at the end to show them how they developed. Sometimes I give general feedback to the class.	Highly	Medium I don't have the knowledge about that. I try to help but not in this precise, organized manner.	Highly Yes, it is connected to the previous one.	Highly	Medium I know the concept out of my experience, but I need technical knowledge about how to do it.	Highly Students have to see this and how they move from one step to the next.
Intervening in the learning process either by taking an	Medium	Medium I am aware of it	Highly	Low No, I don't know about that.	Highly I do it	Highly	Medium In terms of concepts, I don't know them, but	Low

"interventionist" or "interactionist" role						Mainly by experience. I know about them, but I need to practice them.	when you explained them, I do some of them unconsciously and not based on technical input.	I want to know-how and the techniques of doing it. I have never done this.
Scaffolding their learning through different mediating exercises "graduated prompt", "testing the limits", and "mediated learning experience".	<b>Low</b> We don't have this practice	<b>Low</b> I am not very aware of how the task is designed.	<b>Medium</b> I don't use them	<b>Low</b> No, I don't know about this I was never exposed to this.	<b>Low</b> I am not very familiar with them.	<b>Medium</b> Due to time constraint, we cannot apply these.	<b>Low</b> I don't know the concepts, but intuitively I do some of them.	<b>Low</b> This is very new to me.
Using internet resources such as online tutorials and adapt contents to address students' particular needs	<b>Low</b> I never used this	<b>Medium</b> I use some online material	<b>Medium</b> Need more about that	<b>Low</b> I am not used to them	<b>Medium</b> Sometimes I do; we show them to learn from other recourses and themselves.	<b>Medium</b> I have some knowledge of it, but I have not used it.	<b>Medium</b>	<b>Medium</b> I do it
Using assessment methods to monitor, learning by collecting formal data (tests) and informal data (observing) of students' language development.	<b>Highly</b> Experience	<b>Medium</b> As a teacher, yes but as assessor no I just create the exam and convey the result we do not keep track of results	<b>Highly</b>	<b>Low</b> I have never tried that.	<b>Highly</b> I do it as a self-initiation	<b>Highly</b> From my teaching experience	<b>Highly</b> Again, through my teaching experience	<b>Highly</b> By experience
Providing encouraging feedback that is meant to modify the learning process not just a score	<b>Highly</b> Experience	<b>Highly</b> I usually believe in motivating students	<b>Highly</b>	<b>Medium</b> I don't have the theoretical preparation, but I do this out of my ethical perspective.	<b>Highly</b> Yes, I do it but again, not required.	<b>Highly</b> From my teaching experience	<b>Highly</b> Again, through my teaching experience	<b>Highly</b> By experience
Monitoring, recording, and reporting student language development	<b>Highly</b> I can do it, but we don't do it.	<b>Medium</b> I do it by experience, but I guess I need to know how to do it technically.	<b>Highly</b>	<b>Medium</b> I know about that, but I have never practised this in class.	<b>Medium</b> I don't do it the way it should be happening. I am doing it informally.	<b>Highly</b> We have the banner and the grade books, and we can see the development. Form their work I can tell	<b>Highly</b> Again, through my teaching experience	<b>Highly</b> By experience
Evaluating how well learners have succeeded in achieving the final learning goals	<b>Highly</b>	<b>Highly</b> The score reveals a lot in certain cases.	<b>Highly</b>	<b>Medium</b> I don't know techniques or theoretical knowledge, but I can tell by practice through observation.	<b>Highly</b> We do not do it. I know how to do it.	<b>Highly</b> I don't do it, but I know about it. We can do it through keeping portfolios.	<b>Highly</b> I always do this because I prepare assessments based course objective.	<b>Highly</b> By experience
Improving instruction based on assessment results and feedback	<b>Highly</b>	<b>Highly</b> Sometimes based on the results or the outcome, we develop instructions.	<b>Highly</b>	<b>Low</b> I don't have this knowledge.	<b>Highly</b> I do it	<b>Highly</b> From experience	<b>Highly</b> Yes, of course, I know how important and beneficial this is	<b>Highly</b> By experience

Sub Themes	5. Participants' Responses Regarding their Level of Confidence with the Knowledge of Peer & Self Assessments							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Knowledge of benefits of involving students in assessment	<b>Highly</b>	<b>Medium</b>  I do it with writing assignment and presentation. I distribute rubrics. I ask the students to grade each other and themselves.  That is all, but I don't know more.	<b>Highly</b>	<b>Medium</b>  By experience but I was not prepared for it, or I have the knowledge.	<b>Highly</b>  Again, by practice and. We are not required by the organization to do it. After I do a project with them, I ask about their opinion. It is not formal; it is just between them and me. When they give me feedback, I change some stuff accordingly.	<b>Highly</b>  I know about it, and I applied it in certain previous contexts.	<b>Highly</b>	<b>Low</b>  It is not very reliable.
Knowledge of strategies of using self-and peer assessment in different assessment tasks	<b>Medium</b>	<b>Medium</b>  As I said, I just show them the rubric and explain what I mean by each item and ask them to grade themselves or others work. I know about it I believe in it, but I don't know about techniques and how to apply it.	<b>Medium</b>	<b>Low</b>  I don't know the strategies.	<b>Low</b>  I don't know about them that much.	<b>Medium</b>  I know about it	<b>Medium</b>  In terms of intuition, I am in the middle just by experience.	<b>Low</b>  I have to be involved as a teacher. I can't trust students with this. The teacher should control the class from A to Z. I do it for the sake of change, to break the monotony, but it is not reliable.
Training learners to effectively participate in assessments	<b>Medium</b>  I can train them. I can involve them in creating or peer- and self-assessments. In this context, we don't practice it.	<b>Medium</b>  I didn't read or was trained on that. I just feel that I want them to be involved. I am following my intuition.	<b>Highly</b>  I use it to get feedback on my assessments.	<b>Low</b>  I can't do this. I don't even have the knowledge.	<b>Low</b>  I don't have the time or knowledge.	<b>Medium</b>  Experience	<b>Low</b>  I am not that competent. Two or three minutes that I do in class, I cannot count myself as a trainer.	<b>Low</b>  If I don't believe in it, why should I train students?
Using self- and peer-assessments to promote learners' ability to notice their learning problems and gaps in their target learning goals	<b>Highly</b>  I do it in the class but not during formal assessments.	<b>Highly</b>  I am confident about this. I want them to notice their problems in a non-exam environment away from the stress of exams and grades in a learning environment.	<b>Highly</b>	<b>Highly</b>  By experience, I use it a lot in class.	<b>Highly</b>  I do it	<b>Highly</b>  From my teaching experience and theoretical preparation	<b>Highly</b>	<b>Low</b>  Never ever
Providing learners with criteria used in judging their work	<b>Highly</b>  We do that	<b>Highly</b>  That is what I do. I feel the more guided and clear the task is, the	<b>Highly</b>	<b>Low</b>  I don't know how to do it.	<b>Highly</b>	<b>Highly</b>	<b>Highly</b>	<b>Medium</b>  I might give them the rubric, but in the end, I

and make them use it to self or peer assess their work and their peers work.		more productive and beneficial the outcome will be. Sometimes I simplify my own rubrics.			I sometimes give them the rubric that I will use to assess their work later on.	From my teaching experience and theoretical preparation	I do again as part of teaching training and experience.	am the one who would finalize the whole work. So it is double work.
Encouraging learners to produce their own rating criteria in groups and use them to assess one another.	<b>Low</b>  I have not tried it	<b>Low</b>  I haven't tried it. I don't believe they're at a level where they can create rubrics. I can make them create their own quizzes and questions, simple stuff like multiple choice and fill in the gap but to tell you the truth, I can't trust it fully.	<b>Low</b>  I haven't tried this one.	<b>Low</b>  How can I judge if I don't know myself?	<b>Low</b>  I didn't have the chance to apply it in practical life.	<b>Medium</b>  I haven't used it in this context.	<b>Low</b>  I haven't tried it. I think this is a good idea I can try it. A very simple one	<b>Low</b>  No, I don't believe they can
Training learners to use portfolios, diaries, continuous cards, digital audios, video diaries, online blogs where learners are expected to collect samples of work and commentary to monitor their own learning process and keep records of how their work improves and develops	<b>Low</b>  I have never tried it. I have never had a portfolio work.	<b>Medium</b>  I believe in it, but there should be certain circumstances for it to work: the teachers should be willing, the classrooms should not be that big, and the purpose of the assessment should be different because if the assessment is just to collect grades, you will not have time to do such things. When teachers have the upper hand and given a chance to focus on the development of students, then it will work. The problem with our courses is the pace and the time; we have quizzes, exams, projects; it is not easy to cope with all of that and at the same time work on this. Honestly, we have never had speech samples as records, but for writing sometimes we had samples of their work for them to refer to	<b>Highly</b>	<b>Low</b>  I have never tried that. They can notice the improvement. Unfortunately, I do not have the time to do this. Plus I was not trained to do this.	<b>Low</b>  I have tried to use it once, but I couldn't in cooperate it in my work.	<b>Highly</b>  I can do it with hard or soft copies of their work. I did in a previous context, and it was very successful.	<b>Medium</b>  I am not using it in the true sense. I am just collecting the work they have done until the end of the project but without feedback or improvement or monitoring their learning. I haven't done it this way.	<b>Medium</b>  I believe in it, but I need to be trained on how to do it.



Sub Themes	6. Participants' Responses Regarding their Level of Confidence with Assessment Interpretation & Communication							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Interpret data related to test design such as item difficulty and item discrimination	<b>Medium</b> I know about item difficulty.	<b>Medium</b> I know some of them	<b>Medium</b> We check item difficulty	<b>Low</b> I don't know how to do this.	<b>Low</b> Haven't done it	<b>Low</b>	<b>Low</b>	<b>Low</b> I don't know about this
Interpret the concept of reliability in language assessment: (Dependability, Item analysis, Reliability Threat)	<b>Medium</b> I know about item analysis. We did it internally no one asked for it, and we made some changes based on it. I am not familiar with reliability threats.	<b>Low</b> No. I am not aware of it.	<b>Medium</b> I was involved, but not much, I guess I need more about that.	<b>Low</b> I don't know this	<b>Low</b> It was not part of my preparation as a teacher or an assessor.	<b>Low</b> I am not sure of the terms. I don't use them.	<b>Low</b> From my teaching course, theoretically.	<b>Low</b> Not at all. I have never used them.
Interpret validity in language assessments (Construct, Content, Criterion, Consequential Validity, and Validity as argument)	<b>Medium</b> I know about some of them but again based on experience. I was not taught these things.	<b>Low</b> I don't know about it.	<b>Medium</b> I need a lot after you explained the difference between them.	<b>Low</b> Low confident. I don't know this exists in theories. I do things intuitively.	<b>Medium</b> I know some of them. I don't know the consequential and validity as an argument. I do such things, but I don't the terminology.	<b>Medium</b> Again, I guess I know the meaning and the terms, but I need to know how to use them.	<b>Low</b> I am not confident. I can categorize them but in term of formal preparation or how to do them no.	<b>Low</b> I have never been trained.
Interpret data from large-scale tests, namely means, modes, medians, bell curves and has the ability to calculate them	<b>Low</b> We only collect the data, but we do very basic ones and	<b>Low</b> I am not sure if I know about this.	<b>Medium</b> Right now, I am using my excel formulas to give me these results, but I would love to learn more about these things specifically more practice	<b>Low</b> No at all	<b>Low</b>	<b>Low</b>	<b>Low</b> No. I had this subject when I was at university. I am not good at numbers, and it did not mean that much to me, and I didn't focus on it.	<b>Low</b> I don't have knowledge about that.
Infer students' strength and weakness based on collected data to communicate it to different stakeholders	<b>Medium</b> We do it, and we work on it.	<b>Medium</b> We do it based on experience.	<b>Medium</b> I do some calculations.	<b>Low</b> I don't do that we don't keep records.	<b>Low</b> I have never done it	<b>Medium</b> Instituitively	<b>Medium</b> I do them intuitively from the result not in terms of statistic at all.	<b>Medium</b> I can do this but based on my experience.
Calculate item difficulty and item discrimination for close-ended items to obtain reliable scores	<b>Medium</b> I do them in a very basic manner.	<b>Medium</b> Experience and practice	<b>Medium</b>	<b>Low</b>	<b>Low</b>	<b>Low</b>	<b>Low</b>	<b>Low</b>

Investigate facility and discrimination indices statistically	Low	Low	Low	Low	Low	Low	Low	Low
Calculate test/item reliability, calculating inter/intra-rater reliability)	Medium I don't have the background knowledge or experience on how to do it. We do it in a very simple way based on results, average, and item analysis.	Low I am not aware of it	Medium I need more	Low Zero-knowledge	Low I haven't done this	Medium I know about them, but I don't know how to do them. I am not professional in doing this.	Low No unfortunately	Low
Investigate validity using statistical procedures	Low Not very confident because I don't use that much and the institution don't require it that much	Low I am not aware of it.	Medium I need more	Low Zero-knowledge	Low	Medium I know the term, but I don't know how to do them. When it comes to calculation and static, I need help. The previous stuff was related to my teaching experience, but when it comes to assessment-based experience, it is always a problem.	Low No again	Low
Articulate interpretation of norm- and criterion-referenced assessment results to a variety of audience: student, school, directors	Medium Yes, but only criterion-referenced ones or classroom-based ones.	Medium Because I am involved in creating the syllabus, teaching in the classroom, and creating exams, I can communicate and articulate results but only for classroom-based ones.	Medium Being in contact with assessment, I can do for the ones we create	Medium	Medium I know it	Medium We already do that by experience.	Medium Maybe criterion one in a medium level.	Medium Yes, if it is related to providing feedback and teaching experience.
Communicate interpretations using different reporting tools: Report cards, documents, criteria, guidance counselling.	Highly We do them on Moodle as per context requirement.	Highly I do it	Highly	Medium I communicate the score without justification. Just in the form of grades	Highly Our feedback is in the form of grades.	Highly We don't do counselling.	Highly By experience	Highly By Experience
Use software such as Statistical Package to communicate results	Low Basic ones	Low I don't know about that I was not exposed.	Low	Low	Low	Low	Low	Low
Participate in committee or school-wide discussions about reforms in assessment-related issues, curriculum, materials,	Highly	Highly We do with other team leader or head of the department. We sit together to review	Highly	Medium	Medium I have participated in such things	Highly	Highly	Highly Yes, I love this

grading policies, accountability policies, school evaluation and school evaluation.		certain exams, reflect on them, discuss certain matters. We try to have a voice in but not on the college level or the higher management level.				I know how to do it and know its importance, but we don't do it.		
Have the skills to coach students to analyze their own assessment results, track their own learning, communicate about their own learning, and plan next steps.	Medium	Medium Not that much. I ask them to use some rubric to monitor their mistakes and notice them but not in such manner I mean formally with steps and officially with stages, like checking on previous and comparing it to the current and the target. I do it randomly, but I haven't applied it seriously.	Medium	Low I don't have the skills.	Medium Not that much	Medium We don't do it	Low I haven't done this. This is a very good idea. It sounds great. I believe in it.	Low

Sub Themes	7. Participants' Responses Regarding their Level of Confidence with Assessment Ethics Skills							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Knowledge of meaning and implications of critical language testing: power, ethics, and fairness	<b>Low</b>	<b>Low</b> I don't know about this. For me, ethics in assessment is related to plagiarism, cheating, originality, being fair while grading, and standardization for grading, for example, to make sure that it is fair. This is I guess one part of it but not everything.	<b>Medium</b>	<b>Low</b> No, I have not heard of that.	<b>Medium</b>	<b>Low</b> I need to know more	<b>Medium</b> I hear about these things for the first time. Maybe I am familiar with concepts like test ethics or fairness but just as concepts.	<b>Low</b>
Knowledge of how to work towards equity, non-discrimination, inclusion, and social justice	<b>Medium</b> I know things intuitively.	<b>Medium</b> I was not exposed to any preparation or trained. I did it intuitively.	<b>Medium</b> I want to know how to implement them.	<b>Low</b>	<b>Low</b> I have no formal knowledge about it.	<b>Low</b> I need to know more	<b>Low</b> I believe in these things, but I don't know how to work towards these things. I don't know.	<b>Medium</b> By Experience
Knowledge of codes and concepts of professional morality	<b>Medium</b>	<b>Medium</b> Intuitively	<b>Medium</b> I want to know how to practice that.	<b>Low</b> I do it intuitively	<b>Low</b>	<b>Low</b> I need to know more	<b>Low</b> I believe, but I don't know-how.	<b>Medium</b> By Experience
Knowledge legal and ethical responsibilities concerning the use, storage, and dissemination of assessment results	<b>Highly</b>	<b>Highly</b> I know well about security, archiving and storage of exams.	<b>Highly</b>	<b>Medium</b> By practice we learnt that on the job.	<b>Highly</b> I know about them. That is what organizations in the Arab world in general focus on and nothing else. We learnt them by practice.	<b>Highly</b> Yes, per practice	<b>Highly</b> By Practice	<b>Highly</b> By Experience
Observing guidelines for ethics used in teaching contexts in regard to language assessment	<b>Highly</b>	<b>Highly</b> I abide by their context, adopted ethical guidelines.	<b>Highly</b>	<b>Highly</b> I do it intuitively	<b>Highly</b> I know them from my work.	<b>Highly</b>	<b>Highly</b> I do what is required and given to me as per the provided checklists.	<b>Highly</b>
Treating all students or users of language assessment with respect	<b>Highly</b>	<b>Highly</b>	<b>Highly</b>	<b>Highly</b>	<b>Highly</b> This is done intuitively. We are educators	<b>Highly</b>	<b>Highly</b>	<b>Highly</b>

Providing assessment practices that are fair and non-discriminatory through assuring distributive justice	Highly	Highly	Highly	Low I can't claim that I can do it the way it should be done	Highly	Medium Technically no. I know theoretically	Highly I try my best as per my conceptual knowledge based on my teaching experience	Highly By Experience
Avoiding construct-irrelevant controversial, inflammatory offensive or upsetting test material	Highly	Highly intuitively	Highly	Medium By intuition	Highly	Highly	Highly I do it intuitively as a teacher	Highly By Experience
Adopting transparent language assessment practices by informing students of what, how, and why of assessments	Highly	Highly I do it I tell them about everything	Highly	Highly	Highly	Highly I like to be transparent, and I enjoy doing it	Medium I try my best. It is adopted in my teaching context	Highly
Involving other colleagues in assessment decision making and practices	Highly	Highly Unfortunately, they don't take part in the development of assessment. I do believe teachers should be part of it; they are exposed to everything in class; the exam should be the fruit and the result of the participation of everyone	Highly	Highly It is not adopted in our context	Highly	Highly	Medium I try, but there are constraints.	Highly
Implementing democratic practices by accepting and being open to constructive feedback from colleagues	Highly	Highly I am very open	Highly	Highly I practice that intuitively.	Highly	Highly	Highly Yes, I do it based on my personal commitments towards my colleges.	Highly In a professional manner for the sake of development, not for criticizing.
Using a range of assessment approaches and multiple measures that allow students to show their knowledge	Highly We do that.	Highly We are in contact with students we know how to evaluate them.	Highly Students abilities, preferences and needs control the choice of assessment methods.	Medium I know this by experience.	Highly	Highly	Highly Yes, I do it by experience and intuitions.	Highly By Experience
Defining a clear purpose for assessments, develop specifications, evaluate the content and conduct a field test examination	Medium We don't do it this way.	Medium As I told you, our learning outcome is our guide.	Medium	Low I don't know how to do it in a precise professional manner.	Medium	Medium I need validation for my assessment from a professional boy or someone to train me.	Medium I got it by practice and experience as I mentioned before	Low
Using tests scores ethically through providing evidence of fairness via statistical procedures such as validity and reliability	Low We don't do that	Low We don't really check it as much.	Low	Low	Low	Low	Low I need to know more	Low

Clearly and honestly informing the inferences and decisions that derive from scores in assessments.	Highly	Highly	Highly	Highly	Highly	Highly	Medium I know how to do it As much as I can from my own interpretation of my assessments but not from a statistic or conceptual-based or theoretical-based knowledge	Highly
Using assessment results for feedback to influence language learning	Highly	Highly I question my assessment and teaching a lot.	Highly	Medium	Highly	Highly	Medium Self-driven. it is not required from my context	Highly
Evaluating the kind of washback that assessments can have on learning, teaching, curricula, and institutions	Medium We don't go that far.	Medium We do it intuitively. We don't have a process for it.	Medium	Medium By intuition. I cannot evaluate, but I can judge.	Medium	Medium	Medium Based on my practices as a teacher not as an assessor	Low
Judging the consequences (intended or unintended) stemming from assessment in own context through examining "consequential validity."	Low	Low Not really	Low	Low	Low	Medium I need more knowledge and cooperation in the department.	Low	Low
Implementing democratic language assessment practices by giving students the opportunity to share their voices about assessment	Medium	Medium Yes, I ask about their views all the time, their feedback and I take it into consideration.	Medium We do informally	Medium I believe in it, but I don't have sufficient knowledge on how to do it professionally.	Medium I do it informally	Medium I can, but students need to be trained.	Medium Individual effort	Low This is my job, from where are they going to get the knowledge.
Critiquing the impact and power standardized tests can have	Low I am not aware of them.	Low Not really	Low	Low I don't have the knowledge or the expertise to do so I can base it on what I observe in the field. This is a commonly shared knowledge.	Medium Based on my experience yes but not enough	Low I need to read more	Low I can feel them	Low

## Appendix 13: Analysis of Data Collected about the Participants' Need for the Proposed Assessment

### Knowledge Base

Sub Themes	Table 3.3 Participants' Responses Regarding their Need to Disciplinary and Pedagogical Knowledge as Assessors							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Knowledge of Second Language Learning Theories:	<b>Needed</b> Assessments and learning are interrelated. We cannot think of them separately. We teach students, they learn, and we assess. It is like a circle. Of course, teaching methodology is essential, but first, we should focus on how learners learn.	<b>Needed</b> I need it because at the end of the day we assess what the students learn.	<b>Needed</b> You have to know about the process of learning and how it is happening because you are assessing the process.	<b>Needed</b> How do I simply assess students if I do not know how they learn, therefore I need to know more	<b>Needed</b> Assessment is linked to learning. They are both bricks of the same building. if I change the adopted learning theory, the assessment will change accordingly	<b>Not Needed</b> Not so much because I can still tailor questions for the students in my context because sometimes some of these theories do not apply in reality. Creating an assessment mainly requires knowing your students and knowing the context you are teaching in.	<b>Needed</b> The assessment is a reflection of what is done in class. It is a circle because you cannot teach the students in one method and test them or assess them in another.	<b>Not Needed</b> No how would this help me as an assessor I do not believe in this.
Knowledge of Second Language Teaching Theories	<b>Needed</b> As an assessor, it is crucial to have this knowledge I assess as per my adopted teaching method. If I adopt a communicative approach, then my test should be a communicative-based as well. I focus on fluency. I could not test accuracy in an exam. An assessor needs to work on the practicality or mechanism of how to link methods to assessments.	<b>Needed</b> I believe the style of teaching affects the assessment, if you are following communicative techniques you need to give them tasks, you will ignore grammar, and you will not include it in your assessment.	<b>Needed</b> Assessments are a reflection of what we teach. They are related	<b>Needed</b> Yes, needed because it is related to the way we will assess students' performance in class based on how we approved them as teachers.	<b>Needed</b> I assess based on how I teach and how they learn. When you teach communicatively, you will adopt formative assessment, projects, deliverables; they need a parallel approach in assessments. You have task-based assessments and others that are content-based, and therefore, the assessment will differ accordingly.	<b>Needed</b> As an assessor, I may have done better if I had known about them earlier or professionally. Now I know about them, and I am using them.	<b>Needed</b> An assessment should be a reflection of how students learn and how teachers teach. They are interrelated, and assessor cannot survive with such knowledge because the credibility and the validity of those assessments are debatable.	<b>Needed</b> Yes, because it shows how we teach and theretofore how they learnt. As an assessor, it will help me in choosing the type of questions; the exam questions should reflect my teaching.
Knowledge of Learning Strategies	<b>Needed</b> I need it as an assessor to address individual differences. We could test material in different ways. We should not base it only on paperwork. We need to make our assessments more balanced. We cannot have one standard exam for all. It is for ethicality and fairness.	<b>Needed</b> This knowledge would affect the design of the exam.	<b>Needed</b> To approach them differently. Each assessment should be based on a level of complexity. Most students feel anxious when they take a test, so having more straightforward questions build their motivation and confidence to proceed with the exams.	<b>Needed</b> As an assessor, if you do not know their learning strategies, how you will assess them. Learning strategies are affected by contextual, cultural, and behavioural factors. As assessor, should be aware of them	<b>Needed</b> Knowing about learning strategies will affect my choice of certain questions. I need to change my assessment based on how they approach learning.	<b>Needed</b> In a way, yes. It will affect how I choose my assessment items.	<b>Needed</b> When designing exams, I need to know if I am approaching their cognitive skills or social skills. The assessment should reflect these things.	<b>Needed</b> Yes. They can assist me in controlling the level of the test and therefore create items accordingly.

Knowledge of Learning Styles:	<b>Needed</b>  We have to design our assessments based on individual needs and preference. Assessments need to address different learning styles. We are dealing with human beings. We have to give chances to each learner. If they cannot succeed in one component, he/she can succeed in another. We need to give them more individualized changes in assessments. Assessors need to think about them during assessment preparation to come up with fair assessments that address different styles.	<b>Needed</b>  If we know about their styles, we can design different assessments that suit different styles. That is why we have oral presentations, projects, written exams. They need to compensate and balance between alternative assessments, but we cannot just focus on one style.	<b>Needed</b>  I do not believe each learner has a separate learning style. I do not believe in categorizing learners. These work together. Also, it is based on the subject matter as well. However, they are required for an assessor, to use a variety of questions.	<b>Needed</b>  I need to know about them because it will help me in shaping questions or task, but I need to know how I can use them effectively and transfer them from just a knowledge into something practical. I need to know them for assessing or creating a rubric for presentation, for example.	<b>Needed</b>  Again, you have to know your otherwise you are putting them in a tight spot and they cannot move that is why if you are as an assessor you have to be teaching as well You cannot be in the back seat just creating exams and producing exam papers and that is it; you have to have hands of teaching in the classes it affects the way you see the assessment.	<b>Needed</b>  I would have this in consideration when I am designing my assessments. I would consider different questions that address styles, levels, needs, preferences.	<b>Needed</b>  Yes, for the diversity of exam questions because not all learners are visuals, like putting a picture. It will not be enough for all students.	<b>Needed</b>  Yes, because there should be different types of assessments based on the types of learners and different learning tests
Knowledge of Theories of Language Use (Language Models)	<b>Needed</b>  This is the content of the assessment we need to know.  We need to know what we are testing, are we testing lexis, grammar or what	<b>Needed</b>  These are our tools. We cannot work without them.	<b>Needed</b>  Absolutely yes otherwise how are you going to be testing? If you are preparing to become a teacher, you should. I cannot perform my assessment tasks without that knowledge.	<b>Needed</b>  A language teacher and specifically a language assessor must be aware of the complexities of language because they need to stop focusing on separate components when assessing language authentically. They need to go beyond sentence structure or the surface level.	<b>Needed</b>  I have to know them, but I do not have to take them like a bible. I believe learners need to acquire the grammatical competence in communication, but it should not be measured as a separate component of assessment. I need them to be able to express themselves using the language.	<b>Needed</b>  Definitely needed in order to create questions that are not difficult to understand, the type of questions, the way a question is written, the instructions of the questions and the certain focus of an exam, what we after are exactly.	<b>Needed</b>  Those are our tools when designing the exams.	<b>Needed</b>  It is important. Tests are a reflection of what we teach.



Sub Themes	3. Participants' Responses Regarding their Need to Theoretical Knowledge of Assessment Purpose, Content, and Methods							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Knowledge of language testing and assessments history	<b>Not Needed</b> We should focus on what we need and do now because language is a living thing. We need to focus on the situation now because methodologies, content, learning styles were not even analyzed at that time. It was a very teacher-based education.	<b>Needed</b> It could be helpful if I could know more about how things developed and how certain tools were illuminated or replaced. Not everything that is outdated is bad. Some practised are turned against for a while and then advocated in another period.	<b>Not Needed</b> I don't think it is required	<b>Needed</b> A least to be familiar with different types of language testing or assessment and to be critical and to come up with your own philosophy	<b>Not Needed</b> I am not concerned about history, but as an assessor, I can search and read about new trends in assessment, which I can use. I don't need to study it as a history I can reflect on how I was assessed as a student.	<b>Needed</b> I might use something from the past with something from the present to come out with a third way that might satisfy the needs of my students.	<b>Needed</b> It is always good to know how it was in the past because it will polish your current or future practices.	<b>Needed</b> It is important to see how they developed I might use some of the past stuff with some levels of students in other contexts it depends
Knowledge of language testing and assessment philosophies	<b>Needed</b> If I were to critique assessments, I need to know the idea behind them. I need to know if my assessments are matching the claimed philosophy of the teaching context.	<b>Needed</b> I can learn about it for second language learners to make it more specific and more related to my field.	<b>Needed</b> Yes, because you should know why you are testing because there should be some kind of outcome. Also, to give them what they need when it comes to assessment design or just to be critical.	<b>Needed</b> These ones are decided by higher management. We need to know their approach because all of the time, they are detached from teachers. I need to critique what is going on around me.	<b>Needed</b> It would be a good thing to read about them and know the result of adopting them, to know where we are, to whom we belong.	<b>Needed</b> It will assist me to find a certain objective for my assessment. Plus, normally these are done on the managerial level; I need to know their philosophy because they kind of control the assessments. I need to know this because I want to defend my point, critique their point because I am one main element in this education operation.	<b>Needed</b> It tells me why I a designing a test, what am I trying to find out, am I evaluating the learners, my teaching or the content?	<b>Needed</b> Yes, to be able to build an assessment. It will control the purpose and the tools I am using.
Knowledge of the Impact of adopting testing or Assessment Philosophy	<b>Needed</b> I need to know what the effect of a certain assessment approach or philosophy to accept it or refute it. As a teacher, I need to know what I am expecting and the result of what I am using on my students.	<b>Needed</b> It is also important.	<b>Needed</b> I need to know more to be more critical, more proficient.	<b>Needed</b> I need to know more. It helps in trying to stand for certain approaches or standing against others professionally and based on valid knowledge, not on intuitions.	<b>Needed</b>	<b>Needed</b> Of course, because I would know if the philosophies can be adopted or not. In the end, teachers, students, everyone involved will be affected. When we were directed to use easier questions, students' knowledge deteriorated because the aim was helping them with grades and higher scores.	<b>Needed</b> Because the philosophy determines the curriculum and the curriculum determines how to teach, and that determine the testing. We need to know these things.	<b>Needed</b>
Knowledge of Major Assessment Qualities: Authenticity, Interactiveness,	<b>Needed</b> Knowing the qualities can help us judge, critique, give feedback. I believe all teachers need to possess this knowledge even if they are not assessing/designing. They need to	<b>Needed</b> I need to know more because it will change the way we assess. We need to keep them as a checklist for our assessment because the type of questions will change accordingly.	<b>Needed</b> These affect the nature of the test, what the students get from the test, the results, and how you are going to evaluate these	<b>Needed</b> Of course, It is the main knowledge that an assessor should no. It guides the assessment design.	<b>Needed</b> These are the most sensitive ones because these help assessors to assess students fairly. Not any questions written on a paper is an	<b>Needed</b> I need to know more about this knowledge, how to make it fair and involve all the elements you mentioned I need more training on this.	<b>Needed</b> I would like to know because this knowledge should be applied to any assessment. They are really important, and I really want to know how to apply it. I have the sense, but I need the practicality.	<b>Needed</b> This affects fairness. To be fair enough when you create assessments.

Practicality, Washback,  Fairness, Ethicality	receive some training in assessment qualities.		results. It will also direct and influence your assessment design.		assessment. These what qualifies assessor			
	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>
Knowledge of Assessment Types (Traditional/Alternative) (Norm-/Criterion-Referenced) (Summative/Formative) (Externally/Internally Mandated)) Standardized/Classroom)	<p>As a teacher of classroom-based or criterion base, I may say I don't need to know about norm-referenced ones if I'm not involved in creating similar ones, but we need to be aware of them because some of our students are involved in such assessments, so we need to guide them if they need help. As an assessor, I have to check their types of questions and how they are testing out there like IELTS and TOFEL. I need to critique them. For example, they are based on a very short-time assessment, and I don't believe they really reveal the skills in the right way. They create anxiety; they are not individualized at all, and they are trying to put everybody in the same box. But in our context, it is different; our assessments are more individualized and varied.</p> <p>I need to know about them to be armed with tools to face them.</p>	<p>As a second language teacher, I believe we need to know about standardized tests as well as classroom-based assessments because at the end we are trying to get them to a certain level, they should have the level of one who could score 6 in IELTS. Plus, our students sit for these assessments, and they can skip the foundation program if they pass them. Sometimes we adopt some questions that are used in these exams because this is something that the students will need. For example, we adopted some questions from the IELTS test. We are trying to create similar tasks in our assessment that are similar to those ones because we think that they are approved and standard because they are the best practices, maybe we need to need to know more to defend what we do and evaluate what we take.</p>	<p>This is putting everything together. They are related to knowledge of assessment philosophy and purpose. We need to know them all. Over the years, we learnt that we could not base everything on summative assessments. We are used to building everything or judgment on traditional ones. We have to work and base our assessment on the process.</p>	<p>Because you need to be aware of what you are creating or at least judge what you receive even if you are not involved in the design. This is not an option.</p>	<p>I will vary my assessment based on these because they serve different purposes. Like we have to know the difference between standardized and classroom assessments because our students are sometimes exposed to these standardized tests like IELTS and TOFEL, and we need to assist them to know how to deal with the tests, and you cannot give advice if you don't have the knowledge plus such knowledge will help me to critique them.</p>	<p>As an assessor, I need to know about them all, maybe in my current context I am not using standardized-norm referenced ones, but in the previous context, I was using them, I was preparing my students for standardized ones. Even if I am not involved in designing them, but I was preparing and training my students to pass exams like the SAT and TOEFL, and I was using some of the grammar and writing questions to train them. It is important knowledge. They are our competitors, we need to know how each is validating their outcome, and each has its advocates. We need to know to critique and be able to form our own personalized, individualized knowledge on how thing should be assessed.</p>	<p>I need to go into the core of these things because I need to make a difference with those young people I am teaching. Again, I need to know how I can apply them. It should be known.</p>	<p>Yes, this will affect everything to know how to design an assessment.</p>
Knowledge of Assessment Methods:  (Tests, Portfolios, Performance Assessment, Self-assessment, Checklists, Interviews, Questionnaires, Role-plays)	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>
	<p>It is definitely needed. These are our tools. I need more skills in designing them.</p>	<p>They test different aspects; you need to know how students are confident with the language, how they use language in real-life situations with role plays and interviews, and you also need to assess their written standard formal abilities. All of them are important to test language use in real life. They all are important we can't depend only on tests, but using a mixture can tell a lot.</p>	<p>Absolutely, if you are after the idea of alternative assessment and against this idea of the test, you should know them. You need progressive forms.</p>	<p>These are our tools we need to know how to design them; you need to know the objectives of each and when to use them because we need to use a variety of assessments.</p>	<p>I need them for different contents I am teaching.</p>		<p>Teachers should know about tests, but at the end, those are summative ones, but we need to see the process, and if I need to see the process I need to know how to use those. Plus, these varieties of assessments address different learners; not all learners perform well in tests, they can do better in interviews, for example. Again, we need to have alternative assessments out of fairness, but again we need someone to tell us how to perform them, competently. I need someone to tell me how to design and work with portfolios it would be good.</p>	<p>They are my tools, especially if we are into a variety of assessment. It does not have to be a test all the time.</p>

Knowledge of Major steps in Language Test Development : (Test Purpose, Construct Definition, Content Specification, test Specification, Item writing)	Needed  If I was trained about such things, I would have been more confident about them. I could have guided the assessor and teachers more. This knowledge would have been reflected in the production of assessment in a more professional way.	Needed  It is very important it helps you organize your ideas, and once you are clear about it, you can be satisfied with your assessment. I think I need more of this; I believe this is the handbook for an exam designer.	Needed	Needed  It is the responsibility of an assessor to know them. These dictates how to create an exam. Before I thought it was easy to design, but after I saw this checklist no, it is not an easy task at all. Not anyone can create an assessment; this is a responsibly it has to be done professionally. Teachers should be involved in this, but they should have the knowledge. Sometimes the rubric we create doesn't match the objective of the task because we don't have the knowledge to align purpose, objectives and chosen items.	Needed  Again, this is the second criteria upon which assessors should be chosen to create an exam after knowledge of assessment qualities. Even when I see exams that are not mine, I can critique based on this knowledge because it is reflected in the professionalism of writing the exam itself. Assessors should not approach an assessment without this; otherwise, someone should train them.	Needed  Definitely, among all the previous I believe this is very important because it has to do with the design of the assessment the students will have in their hands. I need more of this one. In my current place, I have zero development in the assessments.	Needed  This is mandatory	Needed  I need more if I am designing tests.
Knowledge of Cognitions Taxonomies Required for Different test items or Tasks (Bloom's / Webb's)	Needed  They need to be applied to the teaching process first, so we know that the student reached a certain cognitive stage so we can assess based on them.	Needed  It would be very helpful because it related to item design themselves.	Needed  They are very needed for items construction. As the level increase, they really have to demonstrate the knowledge.	Needed  If we want to address individual difference and different level of thinking of students, this is for fairness.	Needed  Because I have to design my assessment in a way to address different levels and it has to reflect these levels because these what differentiate students	Needed  Yes, of course. This is addressing the different cognitive abilities of the students. I need to design different types of questions, direct, indirect.	Needed  It is definitely necessary because it defines the level of questions in the exam; not all questions should be answered by all students. I want to see who is standing out and who is under the level. Can it analyses, synthesis, create? A tester should know these things when designing exams.	Needed  It will affect the choice of the items you will know how to write questions and keeping mind all levels of students.

Sub Themes	2.B. Participants' Responses Regarding their Need to Skill of Designing Assessments (Setting Purpose, Content and Method) – (Procedural Practical Knowledge)							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Identify and state language assessment Purpose	<b>Needed</b>  I have to know the purpose of my assessment.	<b>Needed</b>  Without setting the right purpose, there is no exam.	<b>Needed</b>	<b>Needed</b>  It is part of the daily teaching requirement. it is the first step	<b>Needed</b>  If you miss the purpose of you don't know why you are doing it, the whole assessment will be meaningless; it will not be valid. Like, read aloud this not a reading test.	<b>Needed</b>  I don't know the purpose of my assessments; how I will design the rest of it and perform the other steps.	<b>Needed</b>  If teachers cannot write or speak clearly about what they will test, it will affect all the other components. This is what I think, out of formality and away from my context, the purpose of my assessment should be to see if my students can communicate or not. This my idea because for me language is for communication	<b>Needed</b>
Write the goals and objectives of instructions and consequently, their assessments.	<b>Needed</b>  We cannot assess them based on something we have not taught. It is not fair.	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>  If you lack this knowledge, it will affect the validity of your assessment.	<b>Needed</b>  You cannot come with the assessment without specifying them.	<b>Needed</b>  The learning outcomes are the goal you are looking for; you want to evaluate through assessment if learners acquired what you aimed for (goal) and it is shown) as a learning outcome in the assessment. It is very important to link objectives to assessment.	<b>Needed</b>  I have an objective to reach in class, and it should be the same for my assessment and should be reflected in my instructions and questions. It is a connected circle.	<b>Needed</b>
Align curriculum objectives, instruction, and assessment	<b>Needed</b>  It is mandatory before exam design	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>  It is a must for even classroom assessment.	<b>Needed</b>  To have assessment make sense, to be valid and reliable. It is one circle, the curriculum, the teaching, the assessment. One thing will lead to the other. Without assessment, will be meaningless.	<b>Needed</b>  This alignment is what creates unity. Assessment is the last stage, and it is assessing what we taught and what the students learnt.	<b>Needed</b>  Definitely, if it is not there, your exam is not serving the purpose.	<b>Needed</b>
Define the language construct(s) an assessment will give information about	<b>Needed</b>  Definitely needed	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>  You need to know what you are looking after, the objective of your assessment.	<b>Needed</b>  If you know the objective is to test a certain skill, there should be items that assess this construct. Without defining the construct, it will be like a random thing.	<b>Needed</b>	<b>Needed</b>  Of course, you cannot just tell students just write; there are well-defined sub-skills that we need to address and assess	<b>Needed</b>
Utilize alternative means of assessments to make decisions based on substantive information	<b>Needed</b>  After all, it is a preparatory program that depends on classroom-based assessment, so we need alternative ones, not just exams.	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>  You need to check performance and not to depend on one. Because one tool cannot give a full image of what you want to check		<b>Needed</b>  To address different learning styles. Plus, tests are criticized for being short, not holistic, depriving learners of other chances to prove or show their competences.	<b>Needed</b>  It applies to different styles of learning and ethicality of assessment. There have to be alternatives. The results sometimes confuse teachers because they depend on one assessment, and the teacher knows the abilities of her /his students. Summative tests are not always reflecting the real level of students. We need different styles of assessment for different abilities.	<b>Needed</b>

Construct test specifications/blueprints to design parallel forms of a test	<b>Needed</b> It is important because it makes assessments more standard and formal.	<b>Needed</b> I believe we should have such a thing because it will standardize our assessments. It will really make a big difference. This is guidance. This is the practical part.	<b>Needed</b>	<b>Needed</b> It is very important. I need this professional practical knowledge. For example, I am against standardized language proficiency exams, they don't match the level of our students still sometimes they are required, but I don't know how to critique them, but I have the intuition that things are not done the right way.	<b>Needed</b> We need to be trained because of this need interaction with people that have the experience you cannot just get it from teaching experience, theory or reading. We need to practice them and receive feedback.	<b>Needed</b> It is needed to document the structure of the exam and as a guide for you later and for others later one.	<b>Needed</b> It saves me the effort of recreating things again and again. I need help with this. I need to be trained professionally. It will organize and standardize our work even throughout the academic year. Again when things are written out, well-settled, agreed upon, I have more confidence and trust in my assessments. I will trust their validity, and I can defend them.	<b>Needed</b>
Write selected-response items such as multiple-choice-, true-false, and matching	<b>Needed</b>	<b>Needed</b> I need more	<b>Needed</b>	<b>Needed</b> I want to know how to formulate and how to use them and for which situations	<b>Needed</b> We all need this we are doing constantly	<b>Needed</b>	<b>Needed</b> Of course, because most of our exams depend on them. According to others multiple choice are the easiest, but I doubt that we need more and I watch other assessment designed by others, and it upsets me because by coincidence and luck or simple techniques students can manage to get the answers. We also should not depend on them only because it could be a matter of luck. we need items that make students produce, production of competence, performance	<b>Needed</b>
Design constructed-response items (for speaking and writing)	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b> I need to know more	<b>Needed</b>
Construct well laid-out and perfectly legible items/tasks	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b> This is important because bad laid-out ones confuse students and affect reliability	<b>Needed</b>	<b>Needed</b> We should know step by step more training on how	<b>Needed</b>
Provide clear, explicit, unambiguous instructions	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b> If you don't give the right instruction, I can't rely on the results	<b>Needed</b> because again it affects the reliability	<b>Needed</b>	<b>Needed</b> Definitely needed because it will affect reliability and results	<b>Needed</b>
Provide examples to make candidates	<b>Needed</b>	<b>Needed</b> I need to know more	<b>Needed</b>	<b>Needed</b> It helps to enhance reliability	<b>Needed</b> Again, it affects reliability. And it provides some sort of feedback	<b>Needed</b>	<b>Needed</b> For the same reason	<b>Needed</b>

familiar with assessment techniques								
Include many items to enhance scoring reliability	Needed	Needed I need to know more	Needed	Needed	Needed	Needed	Needed	Needed
Write test syllabuses to inform test users of test formats where applicable.	Needed To prepare students	Needed This will enhance reliability because they know the answer, but they don't know how to write it because they misunderstood the questions.	Needed	Needed Students should be aware of what they will be assessed. I need to know more about how to do it.	Needed We need it but though professional.	Needed It is very important for students.	Needed The students have the right to be informed of what kind of questions and answers are expected. The grades of first assessments are always low because they don't know what to say or write. They are not familiar with the style, instructions.	Needed
Design assessments that are valid not only in terms of course content but also in course tasks	Needed	Needed They are very important. We badly need it. There should be an expert who had the background knowledge and experience in evaluating and judging this.	Needed		Needed Yes of course for assessment to be authentic	Needed	Needed I cannot use items that are supposed to test the reading skill and claim that I am testing writing, for example. Otherwise, it will affect ethicality, fairness, validity, reliability. It is just giving tests to students to get rid of teaching	Needed
Design assessments that are reliable, authentic, fair, ethical, practical and interactive	Needed	Needed but I don't need the theory I need someone to tell me how	Needed	Needed I need to be trained on this to validate what I am doing	Needed	Needed	Needed Badly needed. I need how the criteria to base my work on it. I need practical learning; it will  complete the picture, fill in the blank, make me confident and satisfied about what I do	Needed
Incorporate technologies in assessing students	Needed	Needed	Needed I have to be aware of it, update my assessment tools and make it closer to whatever they're doing and exposed to. It's part of their culture	Needed	Needed Technology is a must now. It motivated some students. It enhances knowledge. It creates an encouraging atmosphere, but it needs proper knowledge on how to do and the cons and pros of doing it.	Needed It is a must now students know more than us. When they go to practical life, the need this competence	Needed I think students like these things it is part of their life know.	Needed Time has changed; we have students who carry iPads, and take notes on them; it is not a good idea to give them only

			now. It is affecting the words and vocab they are using.					paper and pen tests. If we are talking about fairness, it is fair to give them what they need and prefer.
--	--	--	--	--	--	--	--	---

Sub Themes	4. A. Participants' Responses Regarding their Need to Theoretical Knowledge of Assessment Grading (Declarative Theoretical Knowledge)							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Knowledge of grading rational (norm-, criterion-referenced)	Needed	Needed	Needed	Needed  I need to know more	Needed  It affects how the format of the exam and the items you put. Everything is aligned up together. It will affect the reliability and validity. It has to be dealt with	Needed  It would be better to know more about that.	Needed	Needed
Knowledge of methods that change scores into meaningful composites (points, per cent, grades, proficiency levels).	Needed	Needed	Needed	Needed	Needed  Needed but from professionals and we need the practical part, not the theoretical one. How to implement	Needed	Needed  I would like something more scientific or academic I would like to know what they refer to.	Needed
Knowledge of basic linear scaling, weighting scaling, weighting components, and precision of results).	Needed  At least the basic knowledge	Needed  I need technical training	Needed	Needed	Needed  When we do it to one another, and we all lack the knowledge. One has an idea and just share it; it should not be the case when it comes to these practical, sensitive stuff.	Needed  I need practical knowledge about that I need to know more about that.	Not Needed  In my current situation, I don't think I need that much.	Needed
Knowledge of scoring techniques for objectively-marked testing	Needed  We have to be objective.	Needed  More practice and validation	Needed	Needed  This is important for fairness.	Needed  If I am going to be involved in this, I need it definitely.	Needed	Needed  It is important for assessors to have standard knowledge about this because again, it will be related to ethicality and fairness. Students and teachers put a lot of effort, and the assessment is what evaluates all of this effort and to lose control on it at such stage because of ill knowledge of minor aspects is shameful.	Needed
Knowledge of the nature, purpose, and design of scoring rubrics for subjectively marked tests	Needed  They should be graded twice from two different perspectives. I need structured training.	Needed  More on how to do more standardizations, set of rules to follow	Needed	Needed	Needed  If you are involved in assessments, you have to know about these things.	Needed  I need to know more about it to create more fair rubrics for writing and speaking questions.	Needed  I need more of it	Needed  This is very important to know.
Knowledge of consistency and moderation principles for judgment-based assessments.	Needed	Needed  I need this knowledge because we just give a bonus.	Needed	Needed  This is not an easy job. It is not like other components of assessments that you get from your teaching experiences. It is different	Needed  Again, I am interested in the practical aspect of it how to do it.	Needed  I would like to know.	Needed  It is important because it is not just the evaluation of students; it is the evaluation of the evaluation itself. Plus, if I know the rational and the techniques I could be able to defend my results in front of higher management.	Needed  I need to know about it. At least we need to know what managerial level are doing; otherwise, we might be taking part in unethical stuff.



Sub Themes	3 B. Participants' Responses Regarding their Need to Assessment Grading Skills (Procedural Practical Knowledge)							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Construct scoring schemes that quantify learners' performance into useful information for decisions about learners, classrooms, schools and districts	Needed	Needed	Needed	Needed	Needed  If the system does not require it or is not interested in classifying students based on their grades, then it is not needed and if we are not trying to use the students' results for big changes which is not the case in my teaching context.	Needed	Needed  I need to link it to some solid scientific concepts.	Needed
Design scoring keys with explicit instructions for closed-ended responses (right/wrong, checklist, multipoint methods including rubrics and rating scales).	Needed	Needed	Needed	Needed	Needed  for any assessor	Needed	Needed  We need it and more to base it on real trusted sources	Needed
Design scoring rubrics with explicit instructions for open-ended responses (holistic, analytic, primary trait scoring)	Needed  We need to be more detailed. We assume teachers are grading based on the rubrics, but it includes elements of subjectivity. We need to have some training.	Needed	Needed	Needed  Of course	Needed	Needed	Needed  Again, for the same reason, I want to validate what I am doing	Needed
Calculate consistency of scores/ standard error of measurement to assess scoring reliability for both closed-ended and open-ended responses	Needed  It is more than needed	Needed  I need more about it	Needed	Needed  but not for all types of assessments	Needed  It is needed if you are looking for development of learners especially if you are using standardized tests, but again some of our students are sitting for those tests	Needed  I need practical knowledge on how to do it	Needed  As a professional assessor, I need it to justify or defend my decisions	Needed
Calculate scorer reliability through scorer reliability coefficient to quantify the level of agreement among scorers	Needed	Needed	Needed	Needed  Highly needed	Needed	Needed  I need practical knowledge on how to do it	Needed	Needed

Design training workshops for scorers on acceptable approved responses	Needed	Needed	Needed	Needed	<p>Needed</p> <p>I need someone professional to design them for me</p> <p>I didn't have the formal training to do such work</p>	Needed	<p>Needed</p> <p>We need it because if I am designing an assessment and is used by others, there should be this sort of communication about what accepted and not accepted especially that all sections and classes are evaluated based on the same assessments.</p>	Needed
--	--------	--------	--------	--------	---	--------	--	--------

Sub Themes	4. A. Participants' Responses Regarding their Need to the Theoretical knowledge of Providing Feedback ( Declarative Theoretical Knowledge)							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Knowledge of purposes and principles of feedback	<b>Needed</b>	<b>Needed</b> It will help them improve.	<b>Needed</b>	<b>Needed</b> It is the students' right to know their weakness and strength, and this is the only way they can improve.	<b>Needed</b> If we have the objective for students to learn so, they have to have the feedback, so they have to develop.	<b>Needed</b> It is very important because it will be useful for students.	<b>Needed</b> . Because the aim of our assessment should not be only to evaluate them but to shine them. It is a learning experience, but most of the time, we forget this fact, and we focus only on evaluation. Even when they are studying for the test they are learning; when they are trying to cheat in the test they are learning, maybe they will never forget this cheated information because they fought for it. That is why some exams now are an open book; at least they have hand-on learning.	<b>Needed</b> For the sake of development because we should not only see rank.
Knowledge of different types of feedback (descriptive, evaluative, supportive) with their respective functions	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b> Of course, I need to know more.	<b>Needed</b> I need to know more about this because I am involved in this.	<b>Needed</b> Needed but with less number of students	<b>Needed</b> I can give feedback in a very descriptive way. It depends on them, their needs, preference, the whole situation and how much I know or I need to sharpen my knowledge and skills to be ready for them. I can't put everybody in the same basket. Some of them are pain-based. You give them the pain killer, and that is it. Some of them are achievement-based. When I tell, you achieved 60 % the next time, he will achieve more.	<b>Needed</b> I need to know more.
Knowledge of the strengths and weaknesses of various kinds of feedback (task, process, metacognitive, and self-oriented) in facilitating student learning	<b>Needed</b>	<b>Needed</b> I need to know-how	<b>Needed</b>	<b>Needed</b> Of course, needed	<b>Needed</b> I need more If we are doing things in the right way, I have to learn more about them because I have to know which ones I can use with our students because definitely, they are for different levels of learnings. They are different, and they need different types of feedback.	<b>Needed</b> I need to know what works and what doesn't	<b>Needed</b> I would love to know more about them to use them efficiently, especially if they are provided in a more technical practical manner because I believe a lot in feedback.	<b>Needed</b> They will help a lot. I need to use them according to the different types of students and the different situations and subject matter as well.

Sub Themes	4. B. Participants' Responses Regarding their Need to skills of Providing Formative Feedback ( Procedural Practical Knowledge)							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janet	Talen
Setting educational goals and constructing tasks that will move learners towards them.	Needed	Needed	Needed	Needed	Needed	Needed I need more technical, practical knowledge on how to do it.	Needed Yes, short and achievable goals work better.	Needed
Communicating to learners the criteria for success through different communication modes like telling, showing, and having learners discover	Needed	Needed To give them reason and make them clear and motivated.	Needed	Needed It is their right to know such knowledge.	Needed If they don't know why I am doing this, they will lose interest, and they will be demotivated. if I am not transparent with them, they will not take it seriously, and they will not be engaged	Needed I would be effective and motivated if they know.	Needed Assessors really need that because learners have to know why we require them to do this stuff because they are meaningless for them. If they know, they will assist you in what you are doing. It affects their motivation and their trust in their teachers.	Needed
Showing learners the gap between their current stage of development and the next/final stage of development	Needed	Needed	Needed	Needed	Needed	Needed It helps students 'accomplishment and gives a clear image of everything.	Needed	Needed
Intervening in the learning process either by taking an "interventionist" or "interactionist" role	Needed	Needed	Needed	Needed	Needed If you leave it to them, they will not do it this is part. This is motivating for them to learn.	Needed I general I need to have practical training on how to do it or validation of what I am doing is right. I haven't had any training.	Needed I would love to know more	Needed
Scaffolding their learning through different mediating exercises "graduated promptly", "testing the limits", and "mediated learning experience".	Needed	Needed	Needed	Needed	Needed I would love to know about them.	Needed	Needed I believe as assessors; we should know more about that. We need those polishing techniques.	Needed I will love to know if it helps in improving students and my skills as well.
Using internet resources such as online tutorials and adapt contents to address students' particular needs	Not Needed For now, not but maybe in a different context.	Needed	Needed	Needed	Needed We use Turnitin support for feedback. It is really important.	Not Needed	Needed Imagine we give them unrelated texts that make no sense to them. I need more up to date stuff. Sometimes online tutorials are very helpful. I ask them to compare what I shared with what the tutorial provided.	Needed
Using assessment methods to monitor, learning by collecting formal data (tests) and informal data (observing) of students' language development.	Needed	Needed	Needed	Needed It is important	Needed I need it, but it is not required in my context as	Needed We need practicality	Needed Our students do not test scores they are human beings. We have to know them. We have to	Needed

							observe. We cannot only believe a score and neglect observation.	
Providing encouraging feedback that is meant to modify the learning process not just a score	Needed	Needed Students should not give up, and we should not give up as well.	Needed	Needed I need to make sure that they are learning.	Needed I need the students to get better, and they will not without feedback.	Needed I don't teach for the exams. I teach for life. Scores should not be the end mean of assessments.	Needed Technical training on how do it professionally would be really helpful.	Needed
Monitoring, recording, and reporting student language development	Needed	Needed I need more about this.	Needed	Needed Highly importance	Needed We should keep a record I know, but I need to know-how among the big challenging numbers.	Needed	Needed Definitely for the sake of doing it as a process not haphazardly	Needed
Evaluating how well learners have succeeded in achieving the final learning goals	Needed	Needed	Needed	Needed	Needed	Needed It should be there	Needed	Needed Otherwise why we are doing all of this
Improving instruction based on assessment results and feedback	Needed	Needed	Needed	Needed	Needed Otherwise, everything will be the same, and we will not develop. I change my instructions from one class to the other and from one assessment to the other.	Needed	Needed Assessment is not only for learners. It is also for teachers, management, and the whole system.	Needed

Sub Themes	5. A. Participants' Responses Regarding their Need to Theoretical Knowledge of Peer & Self Assessments ( Declarative Theoretical Knowledge)							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Knowledge of benefits of involving students in assessment	<b>Needed</b>  They should not be excluded. If we talk about peer assessment, they learn more from one another. It is better when they work in groups and in pairs. It is less stressful, and they work with less anxiety.	<b>Needed</b>  It is needed because students will be critical about their own work or others. This will be an eye-opener for them to know.	<b>Needed</b>	<b>Needed</b>  I need to know more	<b>Needed</b>  Students are part of the teaching-learning process. They need to have a say.	<b>Needed</b>  Of course, but the number is a factor and how to manage them.	<b>Needed</b>  It is important for assessor and students. The assessor can see how students are looking at things. Plus, it is a chance for learning from another source other than the teacher. It is also motivating. You give responsibilities, they master their own learning, and teachers are stepping back and making them the owner of their own decisions and learning.	<b>Not Needed</b>  I don't believe in it. It is not always effective. Students use it as a way to go out of the teaching model. They start chatting. The weak students are the ones that are bullied; they leave the work to the competent ones.
Knowledge of strategies of using self-and peer assessment in different assessment tasks	<b>Needed</b>	<b>Needed</b>  I need to know more to involve them in my assessment practices because it gives them a chance to reflect upon their own work; it will help them learn more because sometimes when I grade their paper, they just care about the grade but when they grade theirs or one another they look at mistakes, and they have to justify, so it is not the grade that matters it is the content.	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>  If I know them, it will help me to reach my objectives in a smooth way.	<b>Needed</b>	<b>Needed</b>  I need the techniques and strategies I would like to know more.	<b>Not Needed</b>

Sub Themes	5. B. Participants' Responses Regarding their Need to skills of Peers- and self-Assessments							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Training learners to effectively participate in assessments	<b>Needed</b>  It could be much better and fair.	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>  I need training, and it needs to be practised.	<b>Needed</b>  It empowers you as an assessor and the students. They do have a hand on their learning. It is easier for me. It relieves stress. It shifts the focus form me onto the students.	<b>Needed</b>	<b>Not Needed</b>  This is a critical role
Using self- and peer-assessments to promote learners' ability to notice their learning problems and gaps in their target learning goals	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>  It helps the students, when they do them they learn and pay attention to their mistakes. I can see the improvement when I do the following assessment.	<b>Needed</b>  To make them target their weakness instead of having you pointing at it all the time. They also learn from one another.	<b>Needed</b>  I believe in the power of feedback in assessments.	<b>Not Needed</b>  I don't trust the students. I never trust them.
Providing learners with criteria used in judging their work and make them use it to self or peer assess their work and their peers work.	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>  This is very effective when they understand how they are assessed and the criteria we use they do better.	<b>Needed</b>	<b>Needed</b>	<b>Not Needed</b>  I might do it as an ice breaker, not for learning.
Encouraging learners to produce their own rating criteria in groups and use them to assess one another.	<b>Needed</b>  But students need to be guided first and trained.	<b>Not Needed</b>  Because of their level, they can't do this. This is for teachers but not students.	<b>Needed</b>  I guess I need to know more.	<b>Needed</b>	<b>Not Needed</b>	<b>Needed</b>  In higher level and in specific contexts it might be more needed.	<b>Needed</b>  I need to know more about the techniques. it would be beneficial knowledge for an assessor	<b>Not Needed</b>  They are not up to their level. From where would they get this knowledge?
Training learners to use portfolios, diaries, continuous cards, digital audios, video diaries, online blogs where learners are expected to collect samples of work and commentary to monitor their own learning process and keep records of how their work improves and develops	<b>Needed</b>  It is definitely necessarily	<b>Needed</b>  I encourage to keep records of their work to see the development. They need to feel the difference.  It is very important to track the level of the students, especially with weak ones and every writing they need to check and avoid the mistakes they have done in previous ones.	<b>Needed</b>	<b>Needed</b>  I believe this will be very helpful.	<b>Needed</b>  I need more training on how to do it.  When they see their work, they have critical thinking abilities about their own work; it is not one person who is giving feedback all the time. Their colleagues and their own perspectives help.	<b>Needed</b>  It lessens the stress	<b>Needed</b>  I need to know more about the techniques. it would be beneficial knowledge for an assessor  It could be a solution or an alternative for face to face feedback to save time and to manage classes with a high number of students. It can save the headache of corrections, grading, scoring. It works, it is a technological base, and students love it. This will take	<b>Needed</b>  As a teaching and learning tool but not as an assessment tool

		<p>It would be a very good strategy if we had smaller classrooms and motivated students.</p> <p>The teacher has to have time.</p>					<p>us again to the assessment philosophy or purpose adopted by the context or by the teachers themselves. Sometimes it is the teachers who are holding on to a certain philosophy while the context is open to others.</p>	
--	--	---	--	--	--	--	--	--



Sub Themes	6. A. Participants' Responses Regarding their Need to Theoretical Knowledge of Assessment Interpretation & Communication (Declarative Theoretical Knowledge)							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Interpret data related to test design such as item difficulty and item discrimination	Needed	Needed It gives some kind of feedback on students' performance, the items we use. It can tell whether we can use them again for future assessments or not.	Needed	Needed	Needed	Needed	Needed I find them now beneficial.	Needed
Interpret the concept of reliability in language assessment: (Dependability, Item analysis, Reliability Threat)	Needed Of course, needed but we need some training on them.	Needed I think it will help in auditing and assessment. I need to know about them because after you explained it to me, I believe it will affect the fairness of the results and ethicality of the assessment itself.	Needed If done properly, it is definitely needed by the assessor.	Needed Of course, after you explained.	Needed As an assessor, it is not appropriate to device a tool that is unreliable to make sure that the result did not come by chance. If I don't know these, the test will not make any sense.	Needed I need it for the sake of fairness, for evaluating and critiquing my own exams and other exams. I need to make sure that whatever I give them whether in the form of a score, grade, or whatever reflect their knowledge.	Needed For me, any exam should be dependable, so whatever results I get should mean something. The students, other institutions depend on my exams in term of validity and reliability. But in terms of how to apply them, I need assistance and more about them. How to do it in reality	Needed This is important I need to know more.
Interpret validity in language assessments (Construct, Content, Criterion, Consequential Validity, and Validity as argument)	Needed Of course, needed	Needed Again based on the explanation I received, I believe I need to know more about that.	Needed I need to know more about them.	Needed It is a must knowledge now I know how it is important it is about whether my tool is valid or not. This is a huge threat if it is not addressed professionally. It has to do with fairness. The problem is that we are working in private universities which are managed differently and lots of factors interfere in choosing items and questions, how easy exams are, how could they be managed by students, training students. These are business-oriented; this is also a factor in creating assessments.	Needed Definitely, if the assessment is not valid, whatever derived results will not be valid as well.	Needed I need more about it. It makes me confident because there is support or evidence that what I am doing is right or wrong. I need to learn more about them.	Needed It is very important	Needed I need to judge tests. I feel something is wrong, but I don't know how to judge and critique or how to defend my points.
Interpret data from large-scale tests, namely means, modes, medians, bell curves and has the ability to calculate them	Not Needed I don't need a lot. In my context, I need some basic ones, the average, and comparison	Needed	Needed Even if I am not using standardized tests now, our students are using	Needed All are highly important.	Needed I depend. It is better to know, who knows what I am be involved	Needed It would be useful in another	Needed I know that it is needed but before I was not told how important it is or why I would need it. Exactly like our students if you	Needed

	between groups. It does not match what we do in classroom assessments.	It is important to see and gives a kind of a hint on how to design an exam in the future.	them, and we need to guide them. Plus, we might find ourselves involved in such things later on.		in as a teacher in the future as an assessor I should know.	context. But for this one no.	don't tell them the purpose of what you are teaching, they will not be motivated. Maybe if we knew about them, it might have made sense for me because it helps us to critiques our assessments and others as well for the sake of improvements. We have to know the assessment language to be able to work with it.	
Infer students' strength and weakness based on collected data to communicate it to different stakeholders	Needed	Needed	Needed	Needed	We need to be guided, theoretically and practically.	Needed	Needed	Needed

Sub Themes	6.B. Participants' Responses Regarding their Need to Assessment Interpretation & Communication Skills ( Procedural Practical Knowledge)							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Calculate item difficulty and item discrimination for close-ended items to obtain reliable scores	<b>Needed</b> I need more about them.	<b>Needed</b> We illuminate certain questions and affects future designs.	<b>Needed</b>	<b>Needed</b> This is part of ethicality.	<b>Needed</b>	<b>Needed</b> I need knowledge	<b>Needed</b>	<b>Needed</b>
Investigate facility and discrimination indices statistically	<b>Not Needed</b> I don't need it someone else would do it.	<b>Needed</b> It is important	<b>Needed</b>	<b>Needed</b>	<b>Needed</b> Yes. if we do it, we will know what needs to be changed or developed	<b>Needed</b> I need training	<b>Needed</b>	<b>Needed</b>
Calculate test/item reliability, calculating inter/intra-rater reliability)	<b>Needed</b> I need to know more about it.	<b>Needed</b> If the information provided will point out something for me, I need it.	<b>Needed</b>	<b>Needed</b> I would love to know about it. It tells a lot about the item choice itself.	<b>Needed</b> But this depends on working with other professional assessors in a testing committee or a unit to teach me. I can't work on it by myself. I need to be taught.	<b>Needed</b> I would like to know more about this.	<b>Needed</b> I need someone to train me.	<b>Needed</b>
Investigate validity using statistical procedures	<b>Not Needed</b> I need to know more about it dealing with statistical information, especially if I worked in a different context.	<b>Needed</b> I do need it because again they will provide data we are not currently looking at	<b>Needed</b>	<b>Needed</b> It is very important. It has an impact on the quality of the results.	<b>Needed</b> I need assistance from a system or a team of professional Plus. I don't want to have this knowledge, and I am not using it. I need to have systematic procedures, a team to teach me and guidelines and them it will be needed.	<b>Needed</b> I would like to know more about this.	<b>Needed</b>	<b>Needed</b>
Articulate interpretation of norm- and criterion-referenced assessment results to a variety of audience: student, school, directors	<b>Needed</b> Classroom-based ones	<b>Needed</b> These people have the right to know. You have to link everything together. It is not just a number; It indicates a certain thing. 70 means that you have known this and that. They should know that these results mean something.	<b>Needed</b>	<b>Needed</b> It is the students' right.	<b>Needed</b> We need it for development.	<b>Needed</b>	<b>Needed</b> I need to know more about this.	<b>Needed</b>
Communicate interpretations using different reporting tools: Report cards, documents, criteria, guidance counselling.	<b>Needed</b> We don't need official ones as per my context.	<b>Needed</b> They have the right to know why	<b>Needed</b>	<b>Needed</b> If we are not transparent with the students, they will never learn. They will never develop.	<b>Needed</b> Needed It will make the job a lot easier more valid and reliable, and you can compare thing, and it is related to fairness and ethicality because you have precise data.	<b>Needed</b> All involved needs to get some information about the results.	<b>Needed</b>	<b>Needed</b>
Use software such as Statistical Package to communicate results	<b>Not Needed</b> For classroom-based ones, we don't need it	<b>Needed</b> I think it will make our lives easy and it can convey a lot we	<b>Needed</b>	<b>Needed</b> The use of technology can help a lot.	<b>Needed</b>	<b>Needed</b> It depends. I need more practical knowledge and training or at least	<b>Needed</b>	<b>Needed</b>

	maybe for more standardized.					validation of what we are doing.		
Participate in committee or school-wide discussions about reforms in assessment-related issues, curriculum, materials, grading policies, accountability policies, school evaluation and school evaluation.	<b>Needed</b>	<b>Needed</b>  Believe in it because such discussions help in developing or avoiding certain mistakes. It sharpens your competence as an assessor and makes you more confident because you hear other points of views from other assessors, so you become knowledgeable.	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>  Assessments were never a one-person task. When you discuss thing when you have different inputs from other assessors when you know the weakness and strength, you will have more experience, and it will improve the whole process. It will give me more voice. I will ask for help at least I could say I need more support more training. All of this will be reflected in students' performance.	<b>Needed</b>  I am the assessor and the instructor and the one who has daily interaction with students. My voice needs to be heard.	<b>Needed</b>  We need to do something. We need to change bitter facts. We need to have a voice.	<b>Needed</b>
Have the skills to coach students to analyze their own assessment results, track their own learning, communicate about their own learning, and plan next steps.	<b>Needed</b>	<b>Needed</b>  I need to know-how. I need time, motivated students, support from the administrative level. If they don't appreciate these kinds of things and if you have syllabus and material that you need to cover and material, then it is impossible.	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>  It should be given as a formal assessor when they are tasked with anything related to assessments.	<b>Needed</b>	<b>Needed</b>  All trends in education are going towards getting students to have hands-on learning. It sounds democratic	<b>Needed</b>

Sub Themes	7. A. Participants' Responses Regarding their Confidence Need to Theoretical knowledge of Assessment Ethics ( Declarative Theoretical Knowledge)							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Knowledge of meaning and implications of critical language testing: power, ethics, and fairness	<b>Not Needed</b>  I don't know, and I still prepare. I prefer to do it intuitively. I don't need to study this we have these ethics already.	<b>Needed</b>  I need to know more. An exam should not be unfair to learners.	<b>Needed</b>  Personally, I would love to know more. These are based on new findings. These will empower teachers and shake the throne of tests, give teachers the chance to utter their beliefs because, without those mind-opening, nothing will change. Plus, it for the sake of fairness and I guess equal opportunities.	<b>Needed</b>  It is very important for assessors because it focuses on the values.	<b>Needed</b>  It affects the whole generation, tests control, the future of students. We don't need to categorize students. Language assessment is now a factor in many fields; it should not be used as gatekeepers.	<b>Needed</b>  It will change concepts and ways. I know things should be done in a better way, but I don't know-how.	<b>Needed</b>  I believe that the testing system that I have witnessed and experienced as a student and what I am applying now as an assessor is missing something. I am missing the individual factor that is why I need to know more about fairness and ethics and how to apply them practically in assessments. It is not a one-side operation; it is education, and we are working with people. Involving students, teachers are very important. I would like to know more about it.	<b>Needed</b>  It is very important to enhance assessments ethicality and fairness.
Knowledge of how to work towards equity, non-discrimination, inclusion, and social justice	<b>Not Needed</b>  This is intuitive knowledge.	<b>Needed</b>  I need more because you want the exam to reflect the level of students. You want to be fair in terms of variety of assessment you give. You don't want to disappoint students who believe in themselves. Because of a grade, they might believe they are not good enough. Maybe for my own self-satisfaction as an assessor, it would sharpen this knowledge.	<b>Needed</b>	<b>Needed</b>  Needed for fairness and ethicality	<b>Needed</b>  We have to know about them. They are part of everything.	<b>Needed</b>  I need to know how to implement this in my assessments.	<b>Needed</b>  Yes, I believe, but I need training.	<b>Needed</b>
Knowledge of codes and concepts of professional morality	<b>Not Needed</b>  I know enough	<b>Needed</b>  The same	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>  I need to know more about ethics.	<b>Needed</b>  I need the how part	<b>Needed</b>  I need to be trained.	<b>Needed</b>
Knowledge legal and ethical responsibilities concerning the use, storage, and dissemination of assessment results	<b>Not Needed</b>  It is done based on your experience or your institution. This is too theoretical. It is already set up, and you just follow the rules. It is not my role as a designer.	<b>Needed</b>  It is needed, but it should not be the most important aspect of exams ethicality. Keeping it safe and confidential but what is more important is to secure its validity and its content and its design as a dependable tool for evaluating or judging learners' performance.	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>  They are needed, but they are not everything. They should not be the primary concern.	<b>Needed</b>  It is important for confidentiality.	<b>Needed</b>  The same	<b>Needed</b>

Sub Themes	7. B. Participants' Responses Regarding their Need to Assessment Ethics Skills (Procedural Practical Knowledge)							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
Observing guidelines for ethics used in teaching contexts in regard to language assessment	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed
Treating all students or users of language assessment with respect	Needed	Needed	Needed	Needed	we don't want to lose them	Needed	Needed	Needed
Providing assessment practices that are fair and non-discriminatory through assuring distributive justice	Needed	It is important	Needed	Needed	It is the basis of having assessments	Needed	I would love more practical tips	Needed
Avoiding construct-irrelevant controversial, inflammatory offensive or upsetting test material	Needed	Needed	Needed	Needed	Needed	Needed	Needed	Needed
Adopting transparent language assessment practices by informing students of what, how, and why of assessments	Needed	Students should be aware of the purpose. I am motivated by giving them a purpose for learning	Needed	Highly important	We teach based on objectives if they don't know we won't achieve anything	Needed	Needed	Needed
Involving other colleagues in assessment decision making and practices	Needed	I take this feedback and feed the system and develop the cycle	Needed	They have the right to be involved	It gives a different perspective if they are open-minded. People with no experience or training will not help	No matter how you try you still have your soul in the assessment, so you need their ideas and involvement	Needed	Needed
Implementing democratic practices by accepting and being open to constructive feedback from colleagues	Needed	Without feedback, nobody will change and improve. You can only improve when you admit certain weakness and work on it.	Needed	Needed	I accept, but it should be done with the respect it should not be shown to me as a weakness but in the form of constructive feedback	I make changes based on their feedback	Needed	Needed
Using a range of assessment approaches and multiple measures that allows students to show their knowledge	it is by practice Needed	Part of the ethicality is to use different tools to come up with a valid judgement about our students Needed	Needed	Needed	Needed	Needed	We need this knowledge to be fair and ethical Needed	Needed

Defining a clear purpose for assessments, develop specifications, evaluate the content and conduct a field test examination	<b>Needed</b>	<b>Needed</b> I need more about it	<b>Needed</b>	<b>Needed</b>	<b>Needed</b> We need more about them	<b>Needed</b> I need more particle, more validation	<b>Needed</b>	<b>Needed</b>
Using tests scores ethically through providing evidence of fairness via statistical procedures such as validity and reliability	<b>Needed</b>	<b>Needed</b> I need more	<b>Needed</b>	<b>Needed</b>	<b>Needed</b> I need to know-how	<b>Needed</b> I need more	<b>Needed</b>	<b>Needed</b>
Clearly and honestly informing the inferences and decisions that derive from scores in assessments	<b>Needed</b>	<b>Needed</b> Needed to inform higher management for reforms and major changes	<b>Needed</b>	<b>Needed</b> Highly needed	<b>Needed</b>	<b>Needed</b> It is very important to share your ideas with our context otherwise no improvement	<b>Needed</b>	<b>Needed</b>
Using assessment results for feedback to influence language learning	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>	<b>Needed</b> It is like a circle. It will feed the whole process, instructions, assessment and everything.	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>
Evaluating the kind of washback that assessments can have on learning, teaching, curricula, and institutions	<b>Needed</b> It is a huge project.	<b>Needed</b> It is a circle, and it reflects our teaching.	<b>Needed</b> Assessors only cannot perform this task.	<b>Needed</b>	<b>Needed</b> We need it for development.	<b>Needed</b>	<b>Needed</b>	<b>Needed</b>
Judging the consequences (intended or unintended) stemming from assessment in own context through examining "consequential validity."	<b>Needed</b> It needs coordination with different departments	<b>Needed</b> If we are meant to prepare them for these courses definitely yes	<b>Needed</b>	<b>Needed</b>	<b>Needed</b> I need to know about them	<b>Needed</b> need to know more on how to do it in real life	<b>Needed</b>	<b>Needed</b>
Implementing democratic language assessment practices by giving students the opportunity to share their voices about assessment	<b>Needed</b>	<b>Needed</b> It gives insights and helps in developing tools	<b>Needed</b>	<b>Needed</b>	<b>Needed</b> If you have a say in something, you will be motivated.	<b>Needed</b> Because I need to know and get feedback on my work from them	<b>Needed</b>	<b>Not Needed</b> they can give feedback about the difficulty to a certain extent but not to judge
Critiquing the impact and power standardized tests can have	<b>Needed</b>	<b>Needed</b> IELTS and TOFEL are considered holy and sacred, which should not be the case. When we adopt some of the items we are trying to come up with something similar but who told us this is the right thing. We need to know them	<b>Needed</b>	<b>Needed</b>	<b>Needed</b> These standardized tests are becoming a major power everything is controlled by them if I don't know how to critique them I will be at the end of the line	<b>Needed</b> I need how	<b>Needed</b> I need to know how to apply these	<b>Needed</b>

## **Appendix 14: Analysis of Data Collected about the Participants' Context-Adopted Assessment Practices**

Assessment Practices	Table 3.4 Participants' Report about their Contextual Adopted Assessment Practices							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
<b>Assessment Purpose</b>	We are trying to reach roughly B2 level based on the Common European Framework, which is enough for understanding and studying in the English language so they can continue their studies in their academic majors without difficulties by the end of the program. We assess to see if they can do the requirements of that level. Graded summative assessments throughout the academic year evaluate the achievements of these goals and provide grades. In the end, an accumulative grade is given, and a student would be considered as pass or fail according to a pre-set total percentage (70%). The reason behind assessments is to see if they have reached that level. We are assessing the achievement of specific goals or objectives.	To know their level and what they achieved from the course and where they are to come up with scores to level them.	We have specific learning outcomes which we hope for them to achieve after each assessment. We have the formative and the summative ones. The summative is more of an evaluation base because if they get a certain percentage, they get to move to the next step, which will be level two for the foundation and if they are in level 2, then they get to move to undergrad.	We want to check if students have achieved the required objectives and how much they have achieved throughout the semester.	The purpose behind assessment in our context is just to evaluate and grade students.	Assessment in my context is mainly to evaluate student's level of language knowledge and to check their grades. Moreover, if their grades have a specific limit, they have the right to take courses in university to complete their study. Mainly for evaluations of achievement.	To see if they learned the necessary amount of information.	To check their level, understanding of the content that has been taught, and to see how they improved.
<b>Assessment Source</b>	We are doing them internally.	They are internally designed. Questions are similar to ones they are exposed to in their	They are internally designed. Regarding the	There are two sources of assessment, in-class assessments	They are internally mandated; we design them. I use some internet	They are internally mandated. Some questions are from the question bank	Internally designed and mandated.	They are internally mandated. I create the assessments. I depend on the



		textbooks. Textbooks come with CD that has sample questions and exams. We use some these questions, but most of the time we design our own.	source, we do now started creating a test bank, but it is mainly self-created.	that teachers for their students to check the improvement and their performance, which is the ungraded assignment and the other graded assignments, which is done internally by the management.	resources, books for reading texts, course books, and teacher manual.	accompanying the book. Moreover, some are based on my understanding of language knowledge and the curriculum, and some are internet source. Because I have to abide by the policy of the place I work at and exam design that I should stick to, I try to make it a combination of both.		textbook, internet and sample exam they provided me to design similar ones.
<b>Assessment Methods</b>	We have formal graded assessments in the form of mid-term and final exams, where we assess reading, writing, listening, grammar and vocabulary. We have less formal assessment like tests, quizzes, assignments, projects, individual and group presentations students are basically in their comfort zone, in their classrooms. These are graded, but they take place in a very relaxed atmosphere, so students do not feel anxious about them. They do not have very high percentages, but it is good to make them produce something based on the language they acquired. I think they serve very well in terms of seeing the weaknesses of the students and work on them before the exams.	We have presentations, projects, where they have to submit certain deliverables. We have listening assignments, quizzes in-class assessments where they are given a task, and they have to perform it. We also have those formal exams, midterms and final exam. So we have different tools.	We have a variety of assessments, a mixture of formal and informal ones. We have mid-term and final exams. We also have in-class assignments reading, writing, and grammar) we have quizzes, presentations, projects.	We have projects, presentations, quizzes, graded assignments, and in-class activities in addition to the summative ones. We are using formative and summative ones (mid-term and final exams).	We have a variety of assessment methods. We have summative ones (mid-term /final exams), projects, presentations, assignments, and quizzes.	So mainly, assessments we use are a combination of in-class assessments and online submissions in addition to two significant tests, midterm and final exams.	Projects, quizzes, presentations, assignments, and formal exams like midterms and finals	Formative and summative ones

<b>Assessment Decision Makers</b>	They are the team leaders and the head of the department. These people prepare the exams, and the head of the department reviews the exams, and they are stored and printed without having the instructors look at it.	Usually team leaders and head of the department. So the team leader is the person who creates and design the exam and then the head of department reviews and gives feedback.	We are three separate team leaders, one for each module. We are responsible for preparing tests, assignments, and projects. We do not have time. This is a huge load.	Team leader and head of the department	Anyone who is assigned a team leader role perform assessment tasks. We have to create all assessments for all the other teachers who are teaching the course.	Team leaders who are moderating the module	Team leaders design all assessments and related keys and rubrics	So, it is based on one team or one leader. We are creating them; we do not involve teachers. They are not doing it.
<b>Assessment Design Process</b>	There is no specific process. We just have a checklist that we need to fill regarding duration, covered chapters, number of versions, learning outcomes we are assessing, and pre-set types of questions and items that we have to follow strictly with specific percentages regardless of the course nature or our opinion.	We have learning outcomes, material to be covered, and there is a template of sample examples of how the type of questions are used. So, we use them all this is how it is done. We create the first draft. Then we get feedback and then there is a second draft, more feedback, more modifications and changes and editing. And then until we reach the final product.	It is all about meeting deadlines. No one is concerned with the process. We cannot give the amount of attention we would like to give. Most of the time, we are rushing to complete it for the deadline. We cannot give our utmost attention. We just want to finish, submit it and hope for the best. We are just following a checklist with superficial items. It does not touch at all on how we should design; it determines how the assessment	There is no process. We just design based on a checklist of requirements.	We do not have an exam unit. There is no exam unit responsible for guiding us or reviewing exams or even validating our practices.	I was given a certain template, which I should follow; there is no process like setting the purpose, coming up with a construct that fit the purpose. The only process that we should be following is how we should have only four chapters. This is one of the things I find so hard to apply because it is a language. I mean, we are teaching a language, and this language is like the bricks in a wall. It is all connected and pieces of a puzzle.	It may sound funny we receive an email to start preparing the exams then we just prepare our exams, and we upload them for review and approval, and we have to fill in certain forms like 40 per cent of the exam should focus on this type of questions. We are just filling, yes. After a few days, we receive approval. If there are any changes, we just make the changes. But these changes are not in the content of the exam. But in the layout of the exam.	There are no main criteria for that. We have the syllabus; we come up with questions based on skills taught in class. So, there is no process, we come up with the questions based on our experience as students or tests we have seen before. We get old tests, try to paraphrase questions, we try to restructure them. There are no guidelines. There is no training. There is nothing.

			should look like at the end.					
Assessment of Language Skills	<p><b>Reading</b> skill is assessed in mid-term and final exams and quizzes in the form of reading texts followed by multiple-choice and true/false questions. <b>Writing</b> is assessed in mid-term and final exams, where students are required to write sentences, paragraphs, or essays; also, it is assessed as part of the project, when they submit their paragraphs as group submission in Turnitin for grading. The writing topics are related to the topic they were taught about in class. <b>Speaking</b> is assessed by asking students to prepare presentations and present in class. <b>Listening</b> is assessed as a quiz. It was included in the midterm and the final before, but because of technical problems, we separated it from the main exams. And now we are giving it as a quiz as a listening text followed by multiple-choice, true/false or fill in the gaps questions. <b>Vocabulary</b> is assessed only with multiple-choice questions and word formation. <b>Grammar</b> is assessed with multiple-choice questions plus rewrite and error recognition.</p>	<p><b>Reading</b> skill is assessed through reading texts and close-ended questions. <b>Writing:</b> we usually have sentence level questions. We have paragraphs. We provide them with prompts, picture, helping words. We guide them because it is not mostly the content but the way of doing it. So we do not want them to be consumed in what am I supposed to write? But how do I write it? <b>Speaking:</b> We have presentations and in-class debates. We have group presentations, individual presentations. But we do not have this one to one assessments like one assessor and one person. <b>Listening:</b> We have developed listening assessments. It was only true or false and multiple-choice; we added to fill in the gaps questions, and note-taking,</p>	<p><b>Reading:</b> it is assessed in the mid-term and final exams in the form of a text followed by multiple-choice and true/false questions. It is also assessed as an in-class assignment in a more authentic manner, where they are engaged in task-based activities that assess reading and vocabulary and sometimes incorporate with another skill. <b>Writing:</b> we ask them to produce writing pieces on the sentence and paragraph level. We provide prompts and ask them to write. We have to keep it in a simple form because we do not know what the students know. We are pushed to get more on the basic side. We do have other tasks in writing steps,</p>	<p><b>Reading:</b> through quizzes, exams, tests. Most of the questions are multiple-choice based on vocabulary inside the context, insertion, vocabulary questions.</p> <p><b>Writing:</b> the students are being asked to write an essay about a certain topic.</p> <p><b>Speaking:</b> presentation and project presentations.</p> <p><b>Listening:</b> listening to a text, and then choosing correct answers in for closed-ended questions. That happens once throughout the semester.</p>	<p><b>Reading:</b> a reading text with multiple choice questions assessing comprehension of the text. Most of them are direct questions asking for information from the text. So it does not like really assess scanning, skimming; these like skills of reading.</p> <p><b>Writing:</b> they are given a topic, and they are asked to write a paragraph or an essay if they are at a higher level.</p> <p><b>Speaking:</b> oral presentations.</p> <p><b>Grammar:</b> It is probably the main focus of the course; it is assessed through written assessments in quizzes and exams in class with multiple-choice questions.</p> <p><b>Vocabulary:</b> they are provided with a list of most used words that could help them in writing or speaking, and they were assessed mechanically</p>	<p><b>Reading:</b> through a reading passage with questions. Some of the questions are multiple-choice questions; some go for the unity or the ideas, some go for the details, and very few are inference questions. <b>Writing:</b> it is just writing a paragraph and studying the different forms of paragraphs, whether it is an introduction, a conclusion, or a body paragraph, and of course, each one of them has a different structure. For another course, we have the students write a whole essay and of course, with regard to the previous knowledge of the different structures of paragraphs. <b>Grammar:</b> we have one grammar assignment in addition to a final exam. Moreover, it is pure grammar, very direct questions testing the knowledge of rules, which is like drill and practice.</p>	<p><b>Reading:</b> We give them a reading text and ask questions for understanding the gist of the text and some very easy to find answers for like true or false, multiple choice and the answers are right in front of their eyes. So, nothing implied because we do not have time to teach them reading. <b>For writing:</b> It is handled in a better way than reading. They produce something on the exam. <b>For speaking:</b> presentation. <b>For grammar and vocabulary;</b> we assess knowledge of grammar rules. And lists of vocabulary. We have to abide by the context rules.</p>	<p><b>Reading:</b> providing students with a reading selection and asking comprehension questions. Some questions are direct others are indirect: multiple-choice questions, true-false questions. <b>Writing:</b> It is a writing prompt asking students to elaborate on their ideas. We provide them with a topic, and they have to answer the topic based on personal experience, based on background knowledge. <b>Listening:</b> Listening is one of the most challenging skills. We use listening text with questions and answers such as filling the blanks, true-false, and multiple-choice. <b>Speaking:</b> Presentations, we check students' ability to maintain eye contact, use their body</p>

			<p>like combining with transitions, and we have a vocabulary component based on the vocabulary they are given, they have to select from the box to fill sentences.</p> <p><b>Speaking:</b> we do formal presentations. Nothing else is done in speaking.</p> <p><b>Listening:</b> they have one quiz. The questions are based on taught vocabulary within their listening classes and also based on a script, and then they answer multiple-choice, true-false questions based on that script.</p> <p><b>Grammar &amp; vocabulary:</b> they are assessed during in-class activities and in mid-term and final exams as separate sections because their topics are taught as separate modules</p>		<p>through multiple choice questions and fill in the gaps.</p>		<p>language, posture. The presentation is based on a topic that we provide them with. We ask them questions during the presentation, and they answer it.</p> <p><b>Grammar:</b> we test them in quizzes and exams</p>
--	--	--	--	--	--	--	---

<b>Assessment Administration</b>	The exam unit does this one. We are involved in printing. We submit, and they take care of storage, submission on the day of the exam.	It is very controlled. We create the first draft. Then we get feedback and then there is a second draft, more feedback, more modifications and changes and editing. And then until we reach the final product.	There are guidelines. But this is not the only part of the assessment.	They do have a well-settled policy with guidelines.	They do have guidelines for this one.	It is perfectly done when it comes to administration and instructions; this is perfectly happening in a controlled sequence.	Yeah, that is good. Everybody is doing his/her job.	They take good care of the tests.
<b>Assessment Grading</b>	We design rubrics and answer keys internally for open-ended and closed-ended assessments, respectively. It is not something very standard. We prefer to have something more holistic. Teachers sometimes want to have more detailed rubrics, so we adapt them. They want to count mistakes and penalize students.	We have keys and rubrics. We designed those rubrics and answer keys, we test them, and we get feedback, and we improve them.	We design keys and rubrics. Rubrics are holistic. In the past, they were very detailed. So, we would have instructors counting mistakes, but it was not very practical. Regarding grade allocation, it is generally based on the complexity of the items. However, they are in the exam preparation guidelines pre-set for us. We do follow, so we do not come up with stuff.	For the closed-ended questions, we provide answer keys, and for the open-ended like writing and speaking, we provide rubrics.	For writing, we have rubrics. Some of the rubrics do not make sense, to be honest. Sometimes we are forced to use some of the rubrics that have use of charts and graphs as its components, and this does not apply to our writing, and we have to use these; I have no idea why? So, these are enforced; we have no choice. For reading, we use the answer key with the correct answers. And that is it, nothing else.	We follow rubrics, and they are usually included in the same document to help the student know the standards they should follow.	We have keys. We have rubrics. I mean, we do prepare the rubrics based on the classroom practices. And the other teachers have no say in the preparation of the exam or the preparation of the rubrics.	We use keys and rubrics. For speaking, we create rubrics grading students on maintaining eye contact, body language, posture, tone, intonation, voice, clearness, fluency, pronunciation.
<b>Peer and Self-Assessment</b>	We do not have them formally. Some teachers would provide learners with writing or speaking rubrics	I encourage students. I create simplified rubrics, and I ask students to use them to assess not only	We eventually do, but it does not affect their grade. We do provide checklists for the	No, it is not a requirement or an adopted approach I use it in classes by experience, I do not	No, it is not mandatory. It is just an individual effort. If the teacher is convinced that this a	We have self-reflection as one deliverable of the project after we finish all steps. The final one is self-reflection. But we do	No. We do not adopt this fact. Maybe they do not trust the students	As official results, No, we do not, but as a practice. Okay. As a team builder, sometimes

	and ask them to grade oneself or others.	themselves but one another.	students, especially for the writing component. It is not adopted philosophy; it is an individual effort. I prefer it, and I do provide them with a checklist. I try to get them involved.	have professional training	productive tool is in developing student's levels or their learning, they apply them. It depends on how lucky the students; it depends on the mood of the teachers, and time; they do not have time for it.	not give them the freedom to write anything. We give them a set of very direct questions to which they provide answers. But all provide the same unified answers at the end.	because I think they are too away from this concept. I tried myself, and all their concern was the grade and the name of the student who will grade their paper.	students love to do this. As reliable results, I never do.
<b>Assessment Feedback</b>	After grading exams, teachers take them to class. We assume teachers explain how to manage them in a better way the following time. For writing and speaking, we do not discuss rubrics with them in class so as not to embarrass them; we give it in the form of a grade. If they ask for feedback, they can be given once in the teachers' offices.	We usually ask them if they want any kind of feedback. We avoid giving personal feedback in the classroom with the presence of others to avoid embarrassment. Some students do not like to discuss their grades in the classroom, so we invite them. We provide grades and even going through the exam itself.	For quizzes, we go over them with students, so students can see what they have done. With midterms and final exams, students can come to our office, and we can go over the mid-term if they want to discuss grades. The same with writing and presentation rubric, but there is not a feedback process because they are product-based assessments.	Very little feedback. They receive feedback in the form of scores. They can see their quizzes, but not feedback. It is done in an evaluative manner.	We announce grades and bring papers to class if they want to have a look at their assignments. We can just revise the questions in the classes without showing them their actual assignments. In writing, they can visit teachers' offices to have a look at what they have done. And of course, most of them do not. And for reading, listening, grammar, and vocabulary: Nothing.	Besides the scores generally discussing with the students the major mistakes, showing them some of the very common mistakes which of course they identify and there is also personalized feedback for the students who want to know more after giving these general remarks.	After the exam, to be honest, they do not receive real feedback. Again, we do not have time. They receive scores. If and only if the student is curious about the results, he can always come in and explore the paper. And then we give feedback.	To be honest, this is done. But the results of the feedback is not taken into consideration to build other tests.
<b>Assessment Results Communication</b>	They are announced online, and then the papers go to the class where students can see them. So, it is in the form	They are posted on Banner, and then they are asked if they want to see their papers. They	It is related to the whole policy and purpose of assessment. If	By scores but not justified	Grades are posted on	Moodle system or Banner in the form of scores. Moreover,	We put the results on Banner, and they see it, and that is	We show them the grades; we post them on whichever online system we

	of scores. As for feedback, there is no system for that. It is not controlled; we do not have any control over that. It is done as per teacher.	can come to the office, and they can look at their grade, and they can ask for explanations for different items and feedback.	our purpose is learning and development, ten assessment should be communicated differently apart from scores and grades. If we are focusing on evaluation, then it will be all about grades.			sometimes we showed them the exam.	it. No other communications, no discussion, nothing on them. If the student is curious about his or her learning and they feel she wants to see the paper and discuss it with you	have. That is all that we do.
<b>Assessment Reliability</b>	It is not checked. We analyze item difficulty only intuitively; we go through the questions; we see if there is a problem with any type. So, this is personal effort.	We are not checking them in that manner very formally.	They are not checked in great detail. We have time constraints.	Unfortunately, they are not checked.	Nothing not into statistical stuff. We do not use any of such things.	No, they do not check this.	They are not touching on those. Assessments are haphazardly prepared. No tutorials are given to course moderators about how to test reliability or scorer reliability, these sensitive issues.	No, at all.
<b>Assessment Validity</b>	It is not checked	We are not checking them in that manner very formally.	They are not checked in great detail. We have time constraints.	Unfortunately, they are not checked.	Individual effort. No one checks these things. The supervisor just checks technical mistakes or things related to face validity, and sometimes it is just related to sensitive content that should not be there in the exam.	No, they do not check this.	They are not touching on those issues. Assessments are haphazardly prepared. We prepare them by addressing requirements. So it is done intuitively.	No, at all.

<b>Assessment Fairness and Ethical Procedures</b>	<p>We do not have an exam unit that guides us through fairness. We have exam unit, but it is not about the content of the exam. It is mostly about administrative submission and collection. We do checking ourselves unconsciously especially that we create versions of exams. When I review an exam, I pay attention to difficulty level, length of the exam, alignment with objectives, but nobody checks.</p>	<p>Personally, speaking; I am trying to be as transparent as possible when it comes to creating the rubrics, communicating the rubrics with the students, communicating even types of questions. We are trying to make students aware, and we are trying to make them practice. I try; nobody checks if we do such practices; it is personal effort. Some assessors do not pay attention to this aspect at all.</p>	<p>It is not Checked or paid attention to. Everything is happening intuitively ad by hands-on assessments.</p>	<p>They are not checked.</p>	<p>No, there is nothing related to this; they do not pay attention to these kinds of things.</p>	<p>Nothing.</p>	<p>No, at all. These are issues that are not checked. I mean, there are certain points in the checklists, like some of the questions should be for high achievers. Some of the questions should be for low achievers. But how? Based on what? There is nothing clear.</p>	<p>Not always. Because of tests' structure that we create and impose on teachers.</p>
---	--	---	--	------------------------------	--	-----------------	---	---



## Appendix 15: Analysis of Data Collected about the Participants' Beliefs of their Context-Adopted Assessment Practices

Assessment Practices	Table 3.5 Participants' Beliefs about their Contextual Adopted Assessment Practices							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
<b>Assessment Purpose</b>	The adopted assessment purpose is for judging the achievement of certain goals or pre-set learning outcomes. Some assessments are done throughout the semester to check learning progress even though they are not official graded ones, and they do not provide official feedback. If you ask me I believe in more formative assessments than in summative ones so that we can see the process and where learners are, so we can take action accordingly instead of waiting until the midterm exams, which is in the mid-semester and it could be too	I think it is fair. The purpose of the test should be to make sure that they are learning, which is very important. What is happening is that they are being levelled or put into levels based on the assessment results. Moreover, those levels are reflecting what they learned. So basically, whoever worked hard and learned and attended the classes and followed the rules will hopefully achieve better grades, and that will put him at a different level.	I think it is not challenging enough. The question formats are just based on testing knowledge. It is memorization. The practicality is not really there.	I am with the adopted purpose; it is a must and very necessary. Because I need to check the level of improvement and the performance of my students throughout the semester, and I need to give them fair and objective feedback about the performance	I am against this because it is affecting our practices. We have too many assessments. We do not have enough time for teaching. We teach for assessments so that they achieve good grades. Students do not understand the nature of the assessments or why they are assessed. That is why they do not achieve well. They can get good grades. Yes, but regarding learning, I do not think they are learning the language that much. It is teaching for assessment. Students are	I think that the assessment purpose needs to be different. It needs to be more knowledge-based. I think the purpose should be evaluating their language useability, and the type of items should be completely different from the direct questions that we have now.	This is not correct. Assessments should be incorporated within the learning, and it should be a great help to learning. If only we could evaluate and, you know, change or shape the strategy, the material, whatever according to the learners' needs based on the assessment results, but we are not doing it.	I think it is fine. It is an assessment for both rankings and for checking. I mean, to check whether they achieved the learning objectives and for rankings. So, it is a multipurpose not for one purpose.

	late for some students to catch up after that period. The problem is that formal assessments are still dominating because their percentages are high.				memorizing things put them in the assessments, and from one assessment to another assessment to a third assessment without stopping to understand why this is happening. Our assessment practices are demotivating and overwhelming.			
<b>Assessment Source</b>	I am with internal mandated assessment because we have to have assessments that suit our context and our learners' need. We cannot rely on external resources to assess our students with their different levels, backgrounds, language skills, and definitely unique specific culture. So, I think the more individualized assessment sounds, the better they are because it is a sensitive context.	I think it is fair because I think exams should not be something they see for the first time. I think they should be trained and they should be internally mandated or designed.	We definitely need a better testing bank, for instance, for just one exam, I have to come up with 12 to 13 different reading texts for the different versions and questions, and that is just for one level then I have to do it for the other level, plus the makeup exams. It takes a huge amount of my time.	I believe we need to choose external assessments.	It is not structured; it does not have systematic procedures because it depends on every teacher and what she or he does. Some teachers use internet resources blindly. They are not assessing the objectives we should be assessing. We need to have some sort of unified shared guidelines on resources to use.	Actually, when I try to go outside the box, I am told to give back to into the box to change the questions and make them easier, clearer with less critical ideas and fewer questions that are not so clear.	Yes. It is better to do in-class assessments because these external assessments refer to a very large and, round group of participants. However, our terms are doing this to see if our students can continue their studies in their universities. Assessments then should be done in-house because maybe you can tailor it to the objectives of that program and the learning outcomes.	The material should be designed and created by teacher/assessors, but they need to be trained first on how to source assessment, select material, and design tasks themselves, but I still agree it should be internally mandated.

<p><b>Assessment Methods</b></p>	<p>I am satisfied with the variety in assessment methods we use, but I believe we need to manipulate them in order to make them maybe more authentic to reflect the real competence of students. I am not that satisfied with the presentation as the only tool for assessing speaking skill. The problem that students prepare them and present what they prepared; it is not an authentic performance. I think there must be a speaking exam, where students are given a topic to talk about, and teachers would assess their accuracy and fluency. Some grades are awarded to attendance, which I do not consider it an assessment method; I would not penalize or award the student for attending a class.</p>	<p>I think it is fair because some students prefer to show their work in class in a friendly environment. Some students are better at presentations. Some students are auditory, so with listening assignments, they do better than others. Some students like to work individually in an exam environment, where they have certain questions, and they can answer them on their own. So, I think we are offering them a variety of assessments, which is fair somehow. Maybe with formal exams, the allocated grades should be reduced. They should not decide whether a student will pass or not.</p>	<p>It is not fair to base success on summative exams. We have to see them progress throughout the term. However, at some points, it does seem a bit much for the students. It should be spaced out better. From outside, it looks good that there are a variety of assessments and we are not adopting a monotype. But we are turning them into tests by making learners prepare them like they memorize presentation and we grade them with strictly detailed rubrics and give them a score without feedback, the tool is losing its features and value. I believe it is culture-based. We want to see the progress and what students do to prepare, have enough time to provide feedback and work on that feedback.</p>	<p>Although it sounds that we are using a variety of assessments, we are misusing them because we do not have any training. They are not being used as they should be used. It seems that we are using them all in the same way as a tool for collecting grades. They are just used for evaluating and giving scores, so they are not done in a fair way that shows learners' performance.</p>	<p>Not all of them should be used, to be honest. Not all of them are helping or assessing the students' levels. We cannot use them in a standardized manner for every section or every class or every batch of students. Because not all student can be assessed using presentations. Moreover, this affects the way they perform and their grades. Others have to sit for grammar tests and assignment, which they are not competent at despite being fluent users of the language. Plus, the way assessments are implemented differs from one teacher to the other.</p>	<p>Of course, they are designed with good intentions. They are supposed to be following a certain plan that provides a gradual assessment of students from deliverable 1 to 3, which are all based on each other. However, the thing is, sometimes they are not very effective because students either go to certain specified people who do the work for them or if they are asked to do it in class to guarantee that they do it themselves, they find it a bit hard to conduct it the right way. I am not with tools that assess grammar or writing in a product base manner.</p>	<p>Personally, I am against this structure. I would not like to assess grammar as separate content, but, it is the system. It should be within reading and writing. I am not satisfied with the number of assessments. When we are saying that we are doing student-centred teaching and then we are assessing all the time, this is not student-centred teaching. This is assessment centre teaching because we do not have time to teach. I would like my students to write, not answer subject-verb agreement questions. I want my students to write the first draft and then a second draft. We are assessing for the sake of collecting grades. It is not an ongoing assessment as it is claimed to be. I see it as a video game. They are entering from one door until they exit. They have to collect bonuses from here and there, and at the end, those bonuses saved them, and they just pass. Did they learn, questionable? There are good names for assessment that you would love to hear, but</p>	<p>These methods are good because students have to be assessed with different ways and methods, not only one; however, the way I am not satisfied with the way they are applied and the rubrics used to judge these assessments are not very valid.</p>
----------------------------------	--	---	---	--	---	--	--	---

	Sometimes the assessment methods or tools are not used the way they are meant to; students are graded, for example, in projects on bringing visuals.						they are not used as they should be. They are using it as if it is tested. So, it is a testing culture more than assessment.	
<b>Assessment Decision Makers</b>	They are not completely excluded from the process because they know the format of things, and they know the types of questions like they are given the instructions to mark assessments. I think it would be a better idea to involve them more in the preparation and designing processes because they teach the course and they know their students well. Team leaders also teach, but they are only exposed to one group of students. If teachers are involved, there will be more discussions; assessments will be more	I think more people should be involved. I do not think that just a couple of people should be involved in creating exams. So I think that teachers also have to be included. It should not be the responsibility of one person. Moreover, I think that whoever is assigned such a responsibility has to go to certain workshops, training to make sure that they are using the right tools.	No, that is not right. We prepare assessments and rubrics, grade our assessments, teach, observe other teachers and evaluate them, have guiding meeting with teachers. We would like more teacher involvement, but our current setting does not allow it. I believe that they should cooperate and they should provide material because teachers know better about their students.	That is not fair because teachers should be involved in this process as well. They know what suits students, their needs, and their individual differences.	I am not happy about the fact that there is no valid unit giving guidance, instructions, auditing checking. There should be an exam committee, which provides a clear structure. This structure should be used for all courses to ensure the validity of the exam that is being done, but this is not happening in our context.	I think this is wrong. That was not the case before. Previously we used to have a shared exam devising, but they changed this into just the team leader designing the exam. I cannot guarantee that everyone knows what they are doing because they follow the model.	It is risky, and I do not think it is always trusted because sometimes I see exams just compiled from here and there. There is no thought into it. There are no objectives. There are no desired outcomes. The task or the role is given haphazardly; it can be given to people, you know, after certain years of experience, but still, every one of us, including me, needs a kind of training. So at least for the sake of organizational unity, a type of training is needed. It should not be a checklist, but it should be as clean, clear cut training on how to do things, on how to design things so that like almost all the exams will originate from the same source. This is part of faculty professional development, but it	I am not happy about this practice. It is just about one person creating the test. It is based on what he believes. It should be teamwork. I do believe that teachers have to take part in. I think that they know more about students, they are in contact with students.

	individualized and more customized based on the needs of students. Plus, burden and responsibility will be shared, and the load on team leaders will be less.						should be a strong input. If this cannot be done, then maybe one or two people can just get trained, come and share the experience in the institution. I mean, this is a simpler version of things.	
<b>Assessment Design Process</b>	I am not happy. We do it intuitively. We are not guided; there are no specific criteria that we all follow. We need guidance, training and validation for what we do. It is a huge responsibility.	I do not know if it is the right way of doing it if there are other things to be taken into consideration. That is why I think that we need to be provided by training, but I think everything should be reviewed. The problem is with the feedback we receive, which focus on the layout but not on content or how it was designed. I feel like the feedback should be more like thoughtful focusing on the purpose of the exam and the outcome.	This is my concern. When a normal assessment is designed, major stuff should be checked: purpose, construct, written items, keys, and rubrics. Unfortunately, we do not deal with these details. You do not spend enough time with your creation. We do not have time to go deep.	It is not professional. It is not based on expertise. It is not based on training. It is not based on the necessary qualifications.	The head only checks our assessments. She does not go through objectives that are supposed to be covered and the content of the exam and the face validity. I think she does not have the needed information about this. She did not get proper education about these issues.	I am not happy about this because it is just one person designing the assessment. It is a unified point of view. It would be more fruitful if more than one person help together, but they need to be qualified and trained to do so	That is why I am saying that everybody should get a kind of training on how to design the exam because there is no process like setting the purpose, the content, the specification and nothing of the sort. It is not by done based on levels on or steps. I mean exams are haphazardly prepared because I am sure many people do not put the goals and objectives in front of them. I strongly believe in assurance of learning cycle, we teach, we design the exam based on our objectives, goals, and then we evaluate the exam, the results. What is missing, needed, good not that good?	It is totally unethical because you cannot create a test based on the curriculum only. I have different things I need to take into consideration. I have the teacher, I have the students, and I have the curriculum. I have to find a method or tool that meets all. We need to be taught and trained to perform this skill.
<b>Assessment of Language Skills</b>	<b>For reading,</b> I believe the type of questions should	<b>For reading,</b> we could keep what we do. Although they may look	<b>For reading:</b> I am quite happy with in-class assignments, but not	<b>For reading:</b> I am not happy with it because of the	<b>For reading:</b> students need to practice more with	<b>For reading:</b> I think this is not very effective because	<b>For reading:</b> I am not satisfied. We have to teach and assess real	<b>For reading:</b> if the selection itself is based on

<p>be modified; it should only focus on om closed-ended questions. Unfortunately, the current structure is rigid and enforced on us. We need to focus on their competence in comprehending the text. We can use matching ideas with paragraph. I will change the questions in every exam based on the text itself. We are teaching many reading skills, but we are not assessing them like referencing, insertion, skimming, summarizing, analyzing. There are skills that we do not assess like critical evaluation, analyzing, re-ordering. <b>In writing</b>, I believe it is too harsh on students to evaluate them based on one topic, one paragraph in one exam and to ask them to write a well-developed</p>	<p>superficial, the purpose is beyond that; we are trying to help them to understand something without focusing on each and every word: to skim, to scan because we are trying to have them use certain strategies, and this is something that they will need later on in academics. They might need to go through a whole book to for certain information, so they need these simple skills, But we could include similar tasks to the ones we do in class. Assessment can be done in different ways. It does not have to be very formal and scary. It could be done as a practice in the classroom, but not as a test, where they are required to create something based on comprehension; these ones are more purposeful and reflect real authentic abilities but because of time restriction we do not.</p> <p><b>For writing:</b> I like the writing assessment because you give them everything like content-wise, but we focus more on how they organize their ideas, how they</p>	<p>with exams. I believe they are too basic. They need to be more challenging. In the past, we had sentence insertion items that were removed because it was too challenging. For me, something is e missing because when we read, we read for a reason. Maybe we should give them a purpose to read. You need some kind of response because the current response might be depending on haphazard answers or choices. I cannot consider it a valid tool of showing reading competence. There are others out there. I understand we have deadlines for grading because they have tight deadlines. We do opt for the quicker.</p> <p><b>For Writing:</b> No, again, more can be done more. I would like everything to be done during as an in-class writing assignment. I would prefer providing input through a reading/listening/speakin g task, from which they get ideas, come up with their own ideas. Some critical thinking is happening, and then we</p>	<p>questions themselves. The reading assessment does not reflect the true performance of the students. It does not address individual differences among students. It does not also reflect how much learners have achieved throughout the semester. <b>For writing:</b> I am not satisfied because you cannot judge students' performance throughout the whole semester based on one writing in an exam. It is not fair at all. It does not reflect the real development and improvement and the true level of students. <b>For speaking:</b> it is not fair because pre-prepared presentation does not reflect the true competence of students. <b>For Listening:</b> we cannot evaluate or make inferences on students'</p>	<p>reading. They do not assess reading skills. This is reading for academic purposes; it requires different reading skills other finding the correct answer. <b>For writing:</b> I believe it is not a valid way of assessment. It is not assessing students' level in writing based. I think writing needs a long time for preparation. It is not done as a process in this context. <b>For speaking:</b> presentation is not a speaking assessment it is memorizing and coming to speak in front of the audience. There should be some sort of interviews so that you know that they can speak. <b>For grammar:</b> I am against assessing grammar as a separate component. It should be assessed through</p>	<p>reading comprehension should be part of the writing and should not be asking about the details of the passage as much as asking the students about what they think of the passage, not what the passage of the text indicates. <b>For writing:</b> I am not satisfied especially with the summative one; it is a product-based one not a process-based any more. Now they ask them to write during the exam; they just come up with a piece of writing for me to grade, and that is the end of it; no drafting or feedback is given to students. Writing does not take that much time in assessment because of grammar vocabulary items take much time from the writing time. We have some questions on AP, a style which I think is not very feasible or necessary since they just need to know about it. However, not having a test in the rules of AP style and sometimes</p>	<p>authentic reading skill. <b>For writing:</b> we teach them, we assess them by asking them to write, and they produce in the exam. I am satisfied with it. <b>For speaking:</b> I am not satisfied with the rubric we use to assess them. For example, there is an item for assessing nonverbal presentation, which is not clear. There are scores dedicated to subject knowledge and preparation; how could you assess preparation? It is forbidden to change these things. It is totally holistic just to give students some grades.</p> <p><b>Listening:</b> It is a good idea to assess listening because we are preparing them to function academically in English.</p> <p><b>Grammar and vocabulary:</b> I do not like testing grammar as it is so easy to see if the student gets it when he is writing. The vocabulary is tested through reading in my context. Because in the reading passage, we just underline, bold some words which are easily guessable, which are easily understandable. So we ask about those ones that but still using still</p>	<p>student's level, yes. But sometimes you find a gap between the selection of the text and the students' level. The structure is imposed on us. We are talking about power. So, this is the criteria that we have to abide by it. Sometimes questions work with some students, but they do with others. Differentiation works more than having one standardized test, one test, one type of question with a specific criterion that is imposed on you; a specific type of questions, and a specific length of texts. I believe differentiation is the best solution, I think. <b>For writing:</b> I am satisfied with it because we assess the skill directly. I do not believe in processing writing</p>
---	---	--	---	---	---	---	--

	<p>piece of writing in a very limited time. It is product-focused practice; we are teaching for the test. It should be a process, where we use portfolios for example, but we will face three challenges: mixed level of students in the same class, number of students, time, and most important the context adopted policy. Plus, it will come with its problems: absence, interest in grades from students' side, grading, giving feedback; so maybe we can use with an exam at the end. <b>For speaking;</b> we should not depend only on presentations; we should have a speaking exam. <b>For listening,</b> we always wanted to have some note-taking skills, which is not taught or assessed. It could be combined with writing so they can listen to something</p>	<p>write the topic sentences, how they use the connectors, how they develop. In terms of writing process and feedback, this should be done in class as a teaching process not as an assessment process. <b>For speaking:</b> I truly believe in the fact that students should be able to create a presentation and present to an audience and be able to interact and have presentation skills. This is very important in a foundation program. We give them some time to prepare about a specific topic; so they are not tested on the spot on a certain topic; it would not be fair for that student because he might be very fluent, but he does not have enough information about that topic. <b>For Listening:</b> I am fine with the listening part</p>	<p>give feedback, and then they develop. In mid and final exams, there is not a process. That should not be the case. What we are doing is a product-based assessment. When writing goes through these stages, it is more authentic, and that is all we do in the project. Students research something. They come, they write a paragraph, sometimes again, group works, sometimes individual work, and then the instructor provide feedback. They go back, and they change, they adapt, they add and then submit their work. The student is free to submit their work as many times as they would like. <b>For speaking:</b> I believe it is not adequate for speaking to be assessed only using formal presentations. They prepare it over three to four weeks, and they present it in class. Nothing else is done in speaking. <b>For Listening:</b> I am not satisfied; it is done once a term which is not enough. They were not prepared enough for this test, and it should not be</p>	<p>performance based on just one assessment. It has to be a continuous process.</p>	<p>speaking or writing. <b>For vocabulary:</b> they have to try to guess the meaning of the word. It should be assessed in the context.</p>	<p>errors correction. <b>For Grammar:</b> Definitely not satisfied. Not at the college level. This is not what we supposed to be assessed because it is not of use to them.</p>	<p>multiple choice, or matching.</p>	<p>in assessments. Drafting in is in class but not in a test or an exam. I believe that when you test students' writing skill, it has to be only one draft. Do you want to ask them to write the first, then the second, then the third? And then you only grade them on the last ignoring the first and the second, no. Actually, this is not a good reflection of students' writing skill. I use it as a teaching technique, not as an assessment technique. <b>For Listening:</b> It is not a fair way because sometimes the speaker has a very hard accent, and our students are not native speakers. So, they struggle a lot to understand what the speaker is talking about—as for questions, filling in the blanks might be the easiest path best.</p>
--	---	---	---	---	---	---	--------------------------------------	---

	<p>and write about it, but it is a big change out of hand. <b>For grammar</b>, the system has to change dramatically; I do not see grammar as a separate part. The problem for me is teaching it as a separate module. So I would stop teaching grammar separately. Grammar should be integrated with other skills. <b>For vocabulary</b>, it should be used for comprehensive reasons and assisting reading.</p>		<p>again be based on multiple-choice or true-false items. They should interpret, think and share their views. They can take notes. They can write a summary. Based on their listening, they can do many activities. <b>For Vocabulary and Grammar</b>: I believe they should be incorporated in writing, reading, speaking tasks. We expect to see students produce them in tasks.</p>					<p>However, choosing the right answer means that they need to be totally attentive and really focused. So it is really hard, and I am not happy with it. <b>For speaking</b>: I am satisfied with the presentation even if they memorize and pre-prepare the content, I am OK with that. Teachers here play a great role. They need to give them a model and teach them how, first before you ask the students to do the task. <b>Grammar</b>: You should assess students' grammar. It is very important. Because some people are against testing grammar, I believe that grammar should be taught, should be taught and should be assessed.</p>
<b>Assessment Administration</b>	I do not believe we should be involved	We do not need to be less involved in this	This part is fine.	It is controlled one when it comes to	This is the only component of	I am not satisfied at all because I think that this	No, it is not fair, but it is good. Everything is	When it comes to administration,



	in administrative stuff like exams printing.	administrative level; sometimes, it is just time-consuming.		confidentiality and assessment integrity.	assessment that is guided, controlled and checked.	is not the way things should be done. I am not supposed to be saying that, but I am not happy with that at all. Treating exams as if they are sacred stuff and standardizing it in such a manner. However, it should not be the case because this is a course where they are supposed to learn, and we just check their learning and assist them. I mean, I do not like this kind of power, I feel that I am managed, there is much control.	controlled; in this stage, everything is quite controlled. As soon as you print your exams and hand them in, the system starts to roll back, but before that moment, there is no system, but once you hand in your exams, the system starts. You receive emails of the exam schedules, investigators and then sometimes they are fairly distributed. Sometimes they are not fairly distributed, but at least there is a system. This kind of control is not for classroom assessment? It is for standardized tests, and you are not doing standard. This is the dilemma. classroom assessment is based mainly on objectives you are trying to reach and assess whether the students reach those, and then you go, it is an ongoing assessment	they do care about how the tests are kept safe, but they do not care about the content itself. It is all fake.
<b>Assessment Grading</b>	I think we need to have standardization meetings, where we grade samples of paragraphs together without students' names to agree on certain	I am happy about it. Some teachers are giving us feedback. So we are developing. I think with everything we are doing; we are doing it our own way through our own eyes. There might	Rubric and keys were are using now are better than previous ones, but I am not quite satisfied, they can be better. We need some professional guidance. I have some application concerns. For presentation, for	For writing, I believe the rubric is too general and holistic, but we did so because they required it.	I am not happy about this; specifically, they do not reflect what is happening. You do not discuss those kinds of rubric or sit for grading. We	I design the rubrics. I try to make them as easy as possible and as accessible as possible. However, I wish they were detailed. However, again, because of the big number of students, it is	I am not happy with the rubrics we design; they are not professional. It is just based on our intuition. There is no standardization; they are either too holistic or too detailed. Sometimes they are too detailed to	For speaking, I believe rubrics are important; they are good. For reading: I think there is a problem with the grade allocation, the questions that are really

	criteria that can work based on what we have .	be experts who can guide this more.	example, there should be recording because instructors give grades on the spot, and it is impossible to remember everything.		usually use either old ones that are available or the enforced ones. We have to apply. You do not have a say in them. We have to apply them even if they are not responding to the nature of our courses or our assignments, but we have to. Moreover, the ones that are created by us are just like person effort. We do not sit; we do not negotiate. There is no training for grading, scoring.	very hard to make it so detailed. Moreover, we just tried to keep it simple.	the extent that some of them require counting spelling, grammar, and punctuation mistakes. The rubric requires counting mistakes. Are we fishing mistakes for students?	challenging are given fewer points than other questions that are easy. It is not based on specific criteria simply because there is no background knowledge. For writing, listening and reading, keys should be reconsidered.
<b>Peer and Self-Assessment</b>	In my opinion, it should be definitely part of the assessment. Like, for example, in writing assessments, we have checklists for them to check after they finish writing. Basically, they are not grading themselves, but they are checking if they have completed tasks. So it is kind of a self-assessment as	I heard good feedback on what I do. So I think it is it is successful.	Not very satisfied because students just cannot. Maybe it is their culture. They cannot grasp the fact that they need to re-evaluate their work. They do not want to do it. It should be more incorporated into the grading system itself. Maybe if they were trained to do so, they would be more accustomed to it. Also, some teachers give up because they are consciously pushing, but they do not see anything	I think there should be a kind of policy preceded by training for teachers to perform this in class. We, as teachers should be aware of how to do this because it is important. All students must check their improvement and achievement; it is important for teachers to get feedback from students as well	I believe in its benefits. It makes students see their performance from different perspectives. Learners tend to listen to each other more because they know that they are in the same category. Their peers will not just penalize them for mistakes or deduct grades from them. They	It is there, but it is not appropriately applied.	It should be incorporated. It is something good because judging by the age group they learn from each other more and easier than they learn from the teachers. However, there are cultural barriers; they have difficulty accepting criticism. They take very offensive; it may lead to some kind of harsh discussions in the class. They can accept it from teachers. If it is graded, then it is too serious	I do not believe in it. I think that the teacher is the best person to judge and assess her students. I do not trust the students themselves. They might assess each other on a name base, on a criterion that's not valid and they are not up to that level. What kind of knowledge do they have to assess? Even if you give

	<p>well. However, again, students should be trained; they should know how to look at certain things in their own performance. Some learners are harsh on their peers because they are not trained. When it comes to self-assessment, they are influenced by their background because their main concern in this context is the grades. Even if they know that they do not deserve the grade, they still want to make the grade. Once it comes to assessing themselves, they will all give full points to themselves. They will not be critical. They do not reflect on themselves; they do not have the skill. Teachers/assessors should also have the skills, tools and techniques of doing it.</p>		<p>coming back from the students. It is not working as effective and efficient as I would like to see it.</p>	<p>and to make them involved in their learning. It makes the class interactive. They can also have a say in their assessment.</p>	<p>sometimes negotiate. However, with teachers, they just do not negotiate. The instructor is the owner of the knowledge. They do not like even to discuss it with the instructor. So, emotion wise, motivation wise, it is beneficial. It lowers anxiety. It can be done well if teachers are aware of their strategies.</p>		<p>about being left to the hands of a student. They do not trust it. They are grades-oriented, but they can be trained. When they finish university, they are going out to the workforce. They will have to do these things.</p>	<p>the criteria to them. You cannot give knowledge; they are not up to that level. They might learn from doing this. However, do not ask them to grade each other on this. We can use it as icebreaker or teambuilding. Give them the chance to talk with each other. Maybe we can make them detect mistakes, notice theirs unless teachers want to go over their grades again. Their grading is not reliable.</p>
--	--	--	---	---	---	--	--	--

<b>Assessment Feedback</b>	I think this does not work because it does not bring or take them anywhere. Teachers do not put anything on the writing; they underline certain things. We do not have this error codes any more. It should be a process; there should be drafts. Students should take some sort of feedback if there is a second draft, other than that there is no need for the feedback because students know that this was tested and what they will do is negotiating grades. The purpose of feedback should be learning something from for improvement.	I believe it needs to be more personalized, focusing on the strength and weakness of learners, especially for writing and speaking.	There should be a kind of portfolio adopted throughout the whole semester. We would like to see the progression. We do not see the development. There is some feedback given, but it is not that effective or used efficiently.	It should exist because assessment should not be based only on scores. Moreover, I think this should be like a kind of continuous feedback throughout the semester, kind of guidance and counselling for the students' performance and advising them on their weaknesses and how to improve that and also to focus on their strengths and how to enhance their strengths. So, I think it should be a continuous process that should be implemented throughout. We should not focus only on summative exams or scores or grades.	It is not taking anyone further like to advance. Without feedback, how would they know what they are achieving? There should be time allocated for feedback in every course. They will learn from this. They need to reflect on how they are doing. Moreover, this will give them the time to breathe and think it over. Reflection gives them a step forward. However, if you do not have the time, of there is little time allocated for giving feedback, they would not improve.	No, I am not satisfied. I feel that students need more time to realise their weakness. However, if I had fewer students, I would have been better. I am just giving feedback, and I am not sure whether this kind of feedback is effective for them or they grasped it; if they could implement it later. We are not following up whether the feedback is working in the system again to come up with a more or not.	I hate seeing my students as numbers. They are not numbers to me, but I am telling you, I do not have time because right after this exam I have other tasks. If only I have time and I give them feedback back on all these assessments, exams	I think this is good practice. It is good for students and teachers. However, the question is, are you taking this feedback and consultation to build to improve later on? This is not happening. This never happened.
<b>Assessment Results Communication</b>	I am not satisfied with it. There should be an adopted policy for giving feedback or communicating it.	I think it is OK. I think this is how it has to be done.	On Moodle and Banner. Students can see their papers and ask for more feedback if they feel they need. It is all about grades.	The rubrics should be designed in a way that gives room for justification of the results.	It is score-oriented. They are only working towards the numbers. Students care only are looking at the numbers, and the system is	I think we should not only point out their excellence or weakness through grades. We need to give justification for our actions and their performance.	Horrible. I am just giving them certain numbers. I am enhancing and enforcing and reinforcing that idea of grades, exams. I keep showing them that they are at school. This is like a vicious circle. Because	I think we need to negotiate grades; we need to sit with students. Tell them about all mistakes; we have to work on improving those mistakes. I know it

					reinforcing that. They are not looking at their levels or what they have learned. They are following what is needed for all assignments. They are following that instruction given. Moreover, if they are smart enough, they get the best grades, and that is it. They do not have to understand what has happened.		when you give you give nothing but scores when they ask, they ask for nothing but about scores and results.	is time-consuming. However, if you want the students to approve, moreover, if you want to have better results, this should happen, at least with the students fail. Yes. Not all students.
<b>Assessment Reliability</b>	I am not satisfied. These concepts should be paid attention to. Assessors should be trained to perform this task.	I believe we need some help.	Then the main focus is on the content, and the checklist or whatever guidelines are given. I am not satisfied with this.	We just cannot. We do not have the training, any kinds of training or any kind of previous knowledge whatsoever.	I do not feel comfortable. I do not trust our assessment. No. It does not reflect the real thing.	This is because only one is involved in creating the exam or because there is no academic practical training input. I personally did not get any training.	These are very sensitive issues. They need to be checked, and we need to be trained.	Nobody is checking these kinds of things; they are focused on the shape and structure, not even on the content.
<b>Assessment Validity</b>	Higher management and supervisors should check it and be trained on how to verify it. Assessors' work should not be left to intuitions.	I believe we need some help.	Then the main focus is on the content, and the checklist or whatever guidelines are given. I am not satisfied with this.	We just cannot. We did not have any training, any kind of training or any kind of previous knowledge whatsoever.	They are discussions, no feedback, nothing.	I think it would be better if more than one person put some of their spirits in their assessment. Or maybe technical training for teachers. Practical one. We need to know how to do, but maybe from a professional	These are very sensitive issues. They need to be checked, and we need to be trained.	There should be a department specified with checking this; they need to sit with teachers to check the content.
<b>Assessment Fairness and Ethics</b>	I am not happy about it. I would definitely have someone to give us	More attention should be paid. There should be a standardized practice clear to all stakeholders.	There is much to be done. A lot a lot.	The problem is that we are not trained or prepared for this task. We need this	It is affecting everything; individual differences are not	However, supposing yes and there is an action taken to observe fairness, they do not	No, I do not believe our assessment practices are ethical or fair. We have this issue of	These are not fair assessment practices teacher use them in their

	<p>some ideas on how to ensure fairness and ethics in exams. However, this is an issue or a topic that was not raised. So again, let us stress there is no guidance, no criteria, no policy, nor one well-designed stuff to assure us that we are trying to do right.</p>	<p>There should be auditing and supervisions in terms of assessments' fairness. This is ethicality.</p>		<p>professional technical preparation to defend our practices and to improve them.</p>	<p>addressed. Like in the presentation, there is no consideration for individual differences between the students. From the beginning, we are not categorizing according to their levels. So there are mixed levels in the same class, and they have to take the same assessment. It is not realistic. The structure of the items is not addressing different styles, different views. It is like fixed types of questions, we use them, and that is it.</p>	<p>know-how. They care more about making sure that the test is accessible to most students, which is not my knowledge of exams because some of the exam questions have to be addressing different levels of thinking according to Bloom's taxonomy, and this is what I used to do to do before. They are just quantifying stuff; they after quantity, not quality.</p>	<p>success. The success rate should be high, and in order to show or make this success rate high, we have to do anything with everything. We prepare multiple-choice questions with two choices, which is totally unethical. We prepare matching questions with the question and the answer having matching words. So it is like we are giving the grade because they need the grade on a silver plate. Rubrics are rounded. If you are not addressing different individuals and different talents and different strategies and styles; it's unfair because you are focusing on one thing. If I am not giving them valid test, they are not reliable, and therefore the results are not reliable.</p>	<p>class regardless of the level of students regardless of the background; we are not taking into consideration students' levels, different classes, involving teachers. I mean, validity is not there, reliability is not checked. It is professionally done. Just following certain guidelines, go and do it, and that is it.</p>
<p><b>Final Comment</b></p>	<p>I do not believe these assessments reflect learner's language competence. Midterms should be changed because they do not reflect students' real language skills. We should add a speaking component. When</p>	<p>I cannot say it reflects their competence 100 per cent, but I am sure that we are around 70 per cent reflecting their abilities and their knowledge. I think the problem is the lack of guidelines, especially for validity, reliability, design process, choice of items. We are performing our assessment tasks based</p>	<p>A student can pass easily without being that competent in the language. We are busy with any other stuff. For example, students receive 10 per cent of the final grade for attending. Being physically in class does not show anything about their language competence. The instructor can evaluate a</p>	<p>It does not truly show a fair, true honest reflection of the students' improvements throughout the course, because simply the types of assessment that we are using are only for the sake of getting scores, teachers know that.</p>	<p>It does not reflect any of their abilities. I have excellent students, but when it comes to assessments, they get confused. There is no transparency about the aims or the goals of assessment.</p>	<p>The assessments reflect their knowledge of the rules and grammar, how to produce pre-trained writing. Yes. It does not reflect the language use.</p>	<p>No, I am not satisfied. I mean, I am not satisfied with the project work. I am not satisfied with these big exams like midterms and finals. I have this guilty feeling that we are not giving these assessments to these people just to see, just to learn how far we have come. We are giving these things just to</p>	<p>They are not a good reflection of the students' competence. Some students might be really good at English, and they are proficient, and they have really good English language. However, you get</p>

	<p>it comes to exams, they do not perform that well. Furthermore, that is why I find exams not functioning as they should be. The exams are not reflecting their communicative ability and their skills. Something should be done. This should start with placement exams. These students should have at least three levels. We are dealing with a very mixed level of students. We are just trying to put all of them into the same box.</p>	<p>on intuition, feeling and our teaching experience. I believe that we need more training and an expert eye, something with guidelines just to make sure that we are doing it exactly the way it has to be done.</p>	<p>student based on the degree of participation. Based on what the student is producing. Because the assessment does not have to be a test, it could be just observation. Many weak students pass just by guessing because we are not really testing higher-level competencies in the exams themselves and 90 per cent of our assessment are based on multiple-choice and true/false questions. Even with presentations, they can memorize.</p>	<p>Management knows that and students know that, moreover, for the sake of giving feedback and fair feedback for students. So, it is basically done in a very unprofessional way.</p>	<p>Some students pass the program successfully, and they cannot say a word in English. We should have an exam unit with people who are qualified instructors who do not already have knowledge about the assessments. The unit can give clear guidelines, train teachers, show the way to do things and teach the meaning of different terminology and how to apply them. However, most of the teachers, if you ask them, they do not know the concepts. Not any teacher should be involved in assessments. This is affecting students, and it is not fair for them. These are the assessment-based courses, not teaching-based courses.</p>		<p>label them with a grade pass-fail. The more they pass, the better. Did they deserve? Did they deserve to pass? Questionable. Did we assess effectively? Questionable. Were we able to teach them properly? Questionable. So, this is the issue?</p>	<p>really surprised when a student does really badly on the test and when I investigate the problem and analyze it, I find the problem with the tests. There is a gap between what we teach and what we test. We need to understand the terms. You need to know how to apply them. You need some sort of guidance. I need guidance if I am designing exams.</p>
--	---	---	---	---	--	--	--	---

## **Appendix 16: Analysis of Data Collected about the Participants' Suggestions for Better Assessment Practices**

**Table 3.6 Participants' Suggested Solutions for their Current Assessment Practices**

Assessment Component	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
<b>Assessment Purpose</b>	Once we have these levels, then we will have objectives for each level set clear, so assessments' purpose would be for checking achievement of each level's objectives; if students achieve those objectives and show related learning outcomes in their assessments, then they will move on to the next level with new objectives, material, assessments, depending on the curriculum. This will be a smooth transition from one level to another level. So, it is not an assessment of achievement anymore. It is an assessment of the achievement of preset objectives. I would feel more satisfied compared	I want to make sure that learning took place. I want to make sure that students learn something that their language improved.	We have learning outcomes to check.	I would assess to give continuous feedback, fair feedback for my students. I would not assess for the sake of scores only. So we are talking here about a formative purpose of assessment.	I would change the assessment purpose. The purpose of assessment should be assessing preset learning objectives and expected learning outcomes.	To check their language use and to check how they can apply the knowledge of the language into their daily practice and how they can express their thoughts and their feelings in English.	My assessment purpose will be learning because that is why I should use the assessment for. I get the results, and I put them back into the learning circle. Otherwise, the assessment does not stand for anything alone. It is useless, you know, giving them an exam, let us say midterm exam, and then I see where they went wrong mostly. Then it will help me for the betterment of my teaching. It will be for reinforcement of learning too. If they miss something, if they did not get it fully, then I should put it back into the system and make sure that they get what I am giving them, so my purpose would be learning. By the end of that semester, by the end of that year, I can make sure that at least 70 to 80 per cent of my objectives and goals are met.	We assess students not only to check their level. We have to assess to make sure they have learnt and how to improve what they have not acquired.



	to the system now; with this new one, I would feel motivated to work on it. Because I believe it would change and it would improve the whole thing. The whole system; but it is not easy.							
<b>Assessment Designers and Design Makers</b>	I would form a committee with some experienced instructors, but they have to be trained because people are not trained enough to perform this assessment role. We would go for some type of training sessions, which would be given by experts in assessments. Before we start the assessment task, we should have some workshops for practical training; then we should sit and discuss how to apply what we have learned in our context. We cannot just take everything and use them as they are. We have to adapt it to our context. Other teachers may get	That pre-training should also guide us on who are the stars or who can assess. Those people should be selected and given even more advanced training. And then after that, they can be part of the committee or an exam unit, where they can be involved in creating assessment s. I think it should be very specific, but the other	I would again prefer a separate exam team. The team should be able to get some input, some material from other major instructors. As I said before, they know what they are doing in class, and they know what they have covered. Whatever assessment they would create would be based on learning objectives that the syllabus team created. The teachers can teach and assess, and at any point, the students can sit for this proficiency test whenever they feel that they are ready.	I believe that all teachers should be exposed to a certain amount of assessment training, even if they are not going to be involved in assessment, they should have at least adequate ones because we will focus more on classroom-based ones. As for the summative one, we will have a committee; teachers should be involved with opinions and feedback; they should not be isolated.	There should be a committee of assessors trained and led by an expert in assessment, who should be holding sessions for the teacher, sitting with them assessing their assessments, validating, auditing what is going on. This should be done in order to improve assessment methods. The other teachers should not be involved in assessment; they can be consulted about the levels they see suitable for their students, if they find produces assessment difficult for students to understand; they can like suggest some questions or some types of questions to the	If there are a committee and a unit that has the professional knowledge of designing exams, fine. If not based on the condition that we would have a hand, moderators and all the people who are teaching the same course should collaborate in creating the questions, because this reflects different perspectives. Because when one moderator devises the exam for all students, that reflects part of his instruction and that happens involuntarily. So we need more than one point in devising exams. I would like to involve teachers even if you had a centre or a unit or run by professionals.	I said I would like to train teachers to assess, because I strongly believe that the person who creates the exam should be in the class since this is a classroom assessment, after all. I would never like to have armchair testers sitting there and creating tests. You will never create a good assessment without actually teaching the course in the class. It is impossible. I will be selective. Some teachers have potentials to go for this step more than others based on experience in the class. I believe that all teachers should be involved. You would ask for training for everybody, but they would be busy doing something else such as assessment for learning, ongoing feedback, formative ones towards reaching the objectives. Then you	Assessments should be done professionally by professionals and not by normal people. In order to validate those tests, we need professional people. We need specialized people because there are certain criteria which tests should be based on, what are these criteria? I am not sure if I am aware of all these criteria. Teacher's involvement is very important. I cannot imagine a task that is done by someone that is not involved in the teaching itself as a teacher. I know my students more than anyone

	involved by sharing ideas, suggestions but after receiving training from the committee. At least people will understand why we accept or refuse their ideas; they will not take it personally. Whoever is involved in the assessment should be trained. Teachers should be trained about giving feedback, grading written assignments, grading speaking assignment, being objective, being fair, and preparing assignments because now it will be more individualized.	teachers may be given basics stuff to assist them to implement the decision on assessment s tools. Because maybe not everybody is interested in or has the talent or skill to assess.			person involved in the assessments.		leave this summative one for the senior ones who are trained, who are going to be collecting stuff from the pool; it is the last eye, which is the technical checking.	else. I know their weak points. I know their strengths points. So, I should base my tests on my students, some questions should be avoided. Some questions should be written there. So even if someone else writes the test, the teacher has to go over it. She has to take part and give her feedback, change questions, add questions, remove parts and parts. So, teachers' involvement is really important.
<b>Assessment Source</b>	For the placement exam, I would adopt an external source because we need those general standardized exams to place them into levels. And then for our classroom assessment, we have to prepare them and use our resources and design our tasks. We can use books	I believe that assessment s should be internally mandated. They should be varied. We cannot just depend on one tool. I think some assessment s should be in a relaxed	We would tailor our assessments internally, for reading and listening texts, it can be from online, but they must be adapted. I do support the whole authentic idea, but at some points, it has to be adapted based on student's needs, the capacity of students, and their level, but the questions should be sourced or created according to our context	I would not use external sources. I would base assessments on students' differences. I would create different kinds of assessments based on the student's preferences. I would use a variety of assessments, not only paper-based ones or score-based ones.	The source will be the teachers, and it would be done in-house. I would not go to standardized ones. They are not personalized. They are not tailored to them. Assessments should be tailored. So, in-house assessment done by teachers based on goals based on	I would source my assessment according to how I function in class. If more than one person is participating in devising the exam, each one of them should devise questions that reflect the way the language was handled in the classroom. The exam would be task-based. Moreover, by doing this, we would have	In-class formative assessments done by teachers after receiving adequate training, and a summative one at the end internally mandated designed by selected teachers who receive extra training, they do it by the help of the third eye/expert, who is specialized in assessments.	There should be different sources that teachers themselves use as a good source for assessment. Of course, we can use online sites; we can check previous exams. We can use books that we are teaching from. We can get questions from here and

	<p>as resources. If we leave the midterms and finals behind and if we adopt a new system of assessments, then we will not need any of these. We will focus on the performance of the students. However, when the human factor is involved, things will not be objective. In the end, there should be some benchmark as a proficiency exam. It could be an external or an internal one.</p>	<p>environment in class with certain guidelines, and students can just work freely and produce something. Others can be in front of the audience, where students can gain more confidence. Some assessments can be done in a lab, in a computer lab or in a writing lab where there are different tools. So, I believe that it should be coming from different sources and in different ways.</p>	<p>and objectives. We can have other bodies providing us with an exit exam and placement test. They can come from other institutions because they are already standardized. The other ones would be in-class assessments that we would create.</p>		<p>consultation from an expert.</p>	<p>different views of the same topic, which provides more than one perspective, which is, of course, an ultimate end for better exams.</p>		<p>there. There is not only one source,</p>
<b>Assessment Methods</b>	<p>We should change the type of assessments. I am</p>	<p>We can have quizzes, in-</p>	<p>You will have a proper placement test that levels the students, and</p>	<p>I would use a mixture of both formative and summative. Generally,</p>	<p>I would prefer everything that requires students'</p>	<p>We can adopt formative assessments that have more than</p>	<p>Summative assessments will not work very well with this. Maybe I need it</p>	<p>We can have presentations, portfolios, and</p>

	<p>not for setting a mid-term or final exam, especially for writing and speaking. Learners should be assessed in different ways: task-based assignments, projects, short quizzes. By the end of the course, once they reach this last level before the exam is taken, they have to be trained. They have to know the types of questions. What kind of questions are expected in these exams? I mean, it could be similar to an IELTS exam or a TOEFL exam. So mainly teaching and ongoing assessment that would make sure that we achieved our objectives and at the last stage we can start another process in the last month or something like this, preparing for a standardized assessment.</p>	<p>class activities, presentation s, project deliverables , and academic activities, in addition to formal exams. I think that the exam committee should also be involved in how much weight we give every assessment . Cancelling exams altogether and depending only on in-class ones might be an issue because it is needed to determine who is going to the next level. Now, the weight and the design might help us to be more</p>	<p>then you would have ongoing in-class formative assessments to assess objectives we teach them to take them gradually towards our final intended learning outcome. We can have mid-term and final, with the midterm and final, the students would progress to the next level. So, you would use projects and presentations and portfolios but as a means of ongoing formative assessments for progression. However, to exit, there should be a separate exit proficiency exam. Then students can take it every three months if they want. The assessment methods would be just to show the progress of the students and not for evaluating or for having it as a gatekeeper. Again, we will have levels for students to progress from one level to the next. We could have an exam at the end of each level, which would count towards fifteen per cent of the total grade, but it would not be based on</p>	<p>I would focus on small quizzes, assignments, presentations, and projects.</p>	<p>involvement in tasks not in paper and pen and tests. I would go for portfolios, for writing and speaking because this is done with the facilitation and monitoring of teachers. We can have interviews, and these interviews tailored to different students so that we can have different models depending on the level of the students. Feedback should be taking part there everywhere. For example, if students had some tasks, then you give them feedback. They should be assessed now after addressing the feedback they have got, not before it. So, interviews maybe for speaking, portfolios for writing. They can write summaries of their reading to assess the reading comprehension, so it is an assessment of both reading and writing.</p>	<p>one step in doing the same task. This allows self-correction at least for the writing course. I will not use summative methods for writing; it also depends on the skill you are teaching and assessing. All in all, I would prefer task-based assessments. I could assess reading aurally. The oral part has to be an important section of the reading so that it would be an assessment of reading comprehension and at the same time, an assessment of speaking. It should be mainly an integrated task-based activity or assignment. Listening, I would make them listen to a passage and discuss the topic through different items. I would prefer open-ended questions, and I believe that this could improve the whole atmosphere of the classroom and the behaviour of the students; the students would be more motivated. They need to understand this is a learning experience,</p>	<p>at the end after I finish these formative ones and put them back into learning cycle, I can give them a summative one at the end to see again how it worked, but for the summative ones, I cannot put their results back into the same circle. Maybe I can use their feedback for the next semester. Regarding specific methods or tools, I love projects. I would eradicate mid-term exams, and I would put two projects all through the semester. The projects would be long term ones, and there will be many small tiny tasks that work on different integrated skills. So the students get their hands dirty to find the information, to reach a solution, to analyze what they have found, to note what they have learned. So, I can see every single step of this taxonomy, even I can watch them grow with the tasks that they have with these tiny bits of assessments.</p>	<p>questionnaires. Students might do some surveys, written assessments, oral assessments, and formative, summative, different ways to assess students.</p>
--	---	---	--	--	--	--	---	--

		<p>confident about the results we get basically. To be honest with you, with this culture, you have to have something concrete. The problem with learning, if it is too abstract. I want something concrete to make sure that teaching took place, that there was a record of it, that there is a document that proves it. For example, if they want to move to the next level, there should be a report and a record that says they</p>	<p>memorization and knowledge.</p>			<p>not just studying for assessment.</p>		
--	--	--	------------------------------------	--	--	--	--	--

		have the knowledge, it is like a driving license.						
<b>Assessment Design Process</b>	If you are talking about a radical change, like not having exams and having some task-based assessments, then we do not need to make that. It will not be test-based. We will do different formative assessments throughout the whole semester.	The assessment committee is going to be responsible for the assessment design. Based on the curriculum, syllabus, training, and guidelines they have gone through, they will design those exams. It has to come from someone with the knowledge needed.	We would skip the design process because it is going to be on a smaller scale. Teachers should be granted some kind of freedom. Teachers are not testers here; they indeed should have this knowledge in order to judge what they will come to us. We are preparing students for them, but knowledge of criticizing and knowledge of how it is done is different. This is a huge task. I mean, the needs of the other levels.	I would have to make sure that I have the professional training to be fair with the students. I also have to involve the students in the choice of the basic questions and design of the assessment itself, whether it is graded or ungraded. So, they would be involved. I have to get into consideration students' needs and learning styles, as well.	I will leave the assessment task to the assessment committee. I know that I need to be a teacher, not an assessor. If you are going to be an assessor, you really should have the formal training for this, and it should be done through practicality, not theoretically. You do not want to get involved, but I need to have a say in this. I am not supposed to be doing both at the same. Assessment is not my favourite part, to be honest, because I know that it needs more than this. It needs knowledge. It needs time, and I need to give more time to teaching. I feel like I am strong.	I would leave it to the committee, plus the nature of assessment we will conduct will not require that much of control.	It is not like a normal assessment process. I need to make sure that they are learning something to be able to use them in certain situations. I just need to define the steps, set objective of the task, outcome of the task, design it, collect data, comment on it and go on. For the summative, I do not want one person to do it. All teachers should be involved. There will be a pool. There would be one or two senior instructors trained that received the training picking and choosing the questions from the pool, organizing and then going to the third to check the validity, reliability, the statistics, whatever needed.	I will not need the assessment design process because I will not be designing tests. I would have a committee to design tasks more than a test because I am against using tests in the sense of the word, not the regular formal test we know.
<b>Assessment of Language Skills</b>	For writing, it should be monitored through a portfolio. It should be process	I would want to be aware of assessment	In writing, I would like to see the progress of students starting from forming sentences,	For reading, I would go for integrated task-based assessments. I might ask students to	They should not only be trained to get information. They should have	I would prefer having writing taught and assessed as a process using portfolios for	For reading: it has to start very simple because most students are not good at reading: Trying to make	You can test speaking skills when they talk to each other in

	<p>writing. We can start with sentence structure because we need to take them to step by step until maybe essay writing in the second semester, but we cannot ignore that they cannot use punctuation correctly. They have many fragments. They have many run-ons, long sentences. So we have to deal with these things first from the beginning. The base is writing sentences. We cannot build on something, which is collapsing already. If we have levels, especially the first level should work on sentence structure, we have more time to spend on each level. Start teaching as a process; then we will test writing as a process as well. For reading and listening: I would work on them together; we can have small quizzes like to check different reading and listening so that they</p>	<p>knowledge, theory, practice in the field and based on that I might change my assessments altogether; maybe if I am provided with training and knowledge, I might create something completely different.</p>	<p>giving another topic, turning those sentences into a paragraph. I would like to see it constantly going back and forth between teachers and students, providing feedback and students' works on that. I would use portfolios. For reading, it would be based more on research. For beginners, texts would be provided. We would need some multiple-choice items, true-false items, whatever, but more open-ended questions. You are pushing the students to think for themselves and think critically. Get them to read between the lines, and this is something I would like in the end. We could have regular speaking assignments; there would be a speaking pack, and based on that pack, the assessor, which would be another instructor, would ask questions to the students. It would be progressed to more difficult tasks; it would be based on tasks. We can use reading as an input for a speaking task. I guess the problem is that we are</p>	<p>prepare short presentations on reading texts they read. I would include different genres. I would ask them to perform tasks. I would involve them in the process. For writing, I would guide them throughout the writing process, starting from brainstorming and passing by drafting, feedback, until they reach their final draft for speaking. I can ask students to choose any topic of their preference and give presentations. I would have interviews, pair interviews, group presentations. I would ask them to do video presentations, and maybe to make a kind of conversation outside the classroom with their peers. For listening, it would be a discussion; it does not have to be a written exam.</p>	<p>inference questions; write summaries. It should be related to other skills. It should be task-based related to authentic tasks. This generation is not motivated to read. You have to know their interests and needs from the beginning. The topics should not be fixed in all classes. There is a difference between female classes and male classes. You have to differentiate. Writing should be again for a purpose related to their needs. These should write real tasks. It should happen step by step because they need to see their achievements. They should practice, start producing, take feedback, give drafts until they reach the level they are supposed to reach. They should see the process and the development; all of this should be kept in their portfolio. It would give a great effect. Speaking</p>	<p>drafting and keeping records of their drafts and feedback.</p>	<p>them guess vocabulary, interact with the text; I do not want factual questions; I would prepare reading tasks for them; I will ask them just to produce something, do something. Come up with a piece of writing, for example, based on the text, which they have gone deep into. From a map text, I can go down to process writing or speaking. I can directly go to writing. For writing: I would process writing. We cannot ask students to write an essay all of a sudden. I have to teach them how to write sentences first. So this is my first objective sentence skills, and I will assess these sentences, and there will be many ways to assess them and then come up with a paragraph and then go to essay and then different types of essay; I would focus on different genres. For speaking and listening, I do not believe that I have to assess them as separate things, but I can incorporate them; it has to be in a context through a project. Individual speaking tests are stressful for students. It could be a speaking</p>	<p>class. You can test their speaking skills in reading when they read and explain what they have understood, not necessarily individually. They can have a group discussion. We can have group presentations. We can have individual students providing summaries. Listening also is involved in everything. When the teachers introduce any new topic, they need to check their understanding and listening through confirmation and discussion questions. We can give points. This is listening; it does not have to be a standard test. This lesson the pressure, and it is less challenging, and you can make it easy and the less stressful. It can lower their</p>
--	--	--	---	---	--	---	--	---

	<p>are familiar with what they are supposed to do because they fear long texts. We can vary the type of items that assess those skills; we can ask them to perform authentic tasks based on what they read or listened to similar to real-life ones; they have to do something with the information they extract from any visual or audio text; it will vary the genre and registrar of texts. We can expose them to maps, newspaper, messages on social media, and the kind of required tasks will vary accordingly; we can ask them to write or speak about them, so we are assessing their language use. They can create arguments through writing or speaking. We need to move away from this testing culture to an assessment culture, from being test-oriented, grade-oriented, and score-</p>		<p>isolating teaching from assessment, they are interrelated. You can assess while teaching. For listening, they can take note of a conversation taking place in class.</p>		<p>should be part of every skill. Assessments should be task-based, something given to them to produce. All students need interviews when they start working; they need to know how to hold a meeting, how to argue. They will be assessed on how well they do them. I am against having grammar assessments. It should be assessed as part of other skills. Listening has to be related to something. They should be doing something with what they listen to. They should produce something not just to answer multiple-choice and true-false questions.</p>		<p>task related to their disciplines. I would go for an exit interview, for example. I just do not want them to look at this picture and ask what did you see or give me a presentation that is prepared at home. I can assess it when they are speaking about their reading experience, through observing their speaking skills in class, they can talk about their experience with reviewing their peer work and give an oral report about that. It is a genuine environment. It is not a made-up environment, not a made-up atmosphere. The same with listening, I will not let them listen to a tape and answer questions. This is an integrated task-based project. I can, for example, ask them to take notes while their peers are giving feedback about their written work, this is a listening exercise. A the end of the project they can present written pieces or reports; they will be speaking and listening and cooperating with people; they could be assessed on how they use language with the audience in an authentic</p>	<p>anxiety. For reading, while you are introducing any reading text, you can keep one paragraph for students to analyze. You are getting the students away from two or three paragraphs of long texts. You can include direct and indirect questions. For example, underline some words and refer back to the paragraph and guess the contextual meaning of those words. We can use direct questions, critical thinking ones. You can also ask them to summarise this paragraph by two or three sentences if they know how to paraphrase as well. Testing the students reading skill is not always by providing them with reading texts, four to five</p>
--	--	--	---	--	--	--	---	---



	oriented to being performance-oriented. For grammar and vocabulary: They should not be tested separately, they should be tested with the other skills. I do not say that they are not important, of course. They are the most important part of the language, but what we should assess how they are used.						format. It could be a good chance to make them present for their coming instructors whom they will know later. We could invite the business group for students who are going to be specialized in business, and we present the project for them, and they can interact and start talking to the coming professors. We can get feedback from the professors as well and maybe from the feedback of the coming professors we might find if the students are able to move up or not.	multiple-choice questions, true or false, fill in the blanks, matching.  For writing: I think students struggled a lot with it, I think because of their English. Sometimes they write sentences in Arabic if they are not able to express their thoughts in English. So here we need to work more on enhancing their English skills and developing them, starting from the sentence level.
<b>Assessment Administration</b>	We will cancel all of this if we are not going to be using final and mid-term exams. Any we will base all our program on classroom assessments.	I think the current one is good and it is done in the right way, I think the administrative part is also important because you want to make sure that the exam is not shared, the exam is	No, because the impact of every assignment is small. We need to eliminate those kinds of things because we will have the placement and the proficiency exam, which will be coming from somewhere else. Maybe they could be stored.	I do not like to feel detached from anything. I would love to be involved in printing or editing or drafting. We should be involved; we are the teachers and the assessors at the same time. I need to have a say in everything.	Teachers should not be involved at all. It is not their job.		I will only have one big summative assessment; I will not have that much of the job, I guess. Printing, storing, arranging, invigilation once a semester, and how much effort we are saving. We are saving teachers' valuable time that is wasted on bits and pieces of quizzes and grading to the extent that we do not have time for class preparation. The other formative ones do not need that much administrative work; they	

		secured, the exam is done in the right way. I think to a certain degree, that part is also important.					will be designed, prepared and administrated by teachers in class. They are not going to be graded. Maybe I can have coordinators We will plan it in advance weekly. We need to be sure that we are done with goal number one in a week and we will assess it, and then we will give feedback. The feedback will take one week, and then we will go for peer assessment. So, there must be one person to check on.	
<b>Assessment Grading</b>	We are now talking about grading tasks, not exams and tests. We have to have simple rubrics with no many overwhelming items; teachers will be trained on how to design them in advance. Teachers can use error codes and feedback during grading. Rubrics should be uploaded online, where learners are submitting their tasks.	The same with the keys in the rubric, you need some assistance with them because so far you are doing it intuitively based on your teaching experience.	They will already be prepared because they are coming from the outside with the others anyway. I will not bother.	I need to be trained to do it.	The committee would take care of this; they need to be designed based on criteria, not haphazardly.	The rubrics are very flexible material. Then I would create according to these standards and the categories that indeed to fulfil. So rubrics can be modified to any shape the instructor needs depending on the objectives and the learning outcomes.	I will not bother with grading and rubrics; it will be given in the form of feedback to use it to refine their work. I can maybe, for example, prepare the rubric one week ahead, share the rubric with the other instructors. Moreover, I can see their rubrics; what they have come up with, and we can just compile one unified one. You do not need complication and do not want to go as far as grammar and commas. No, I do not want them to fish for mistakes, to fish for punctuation. Maybe if they need extra help with	I always would love to have professional training to improve myself. I might struggle with creating rubrics, especially for listening.

							this mechanism, what about tutoring and writing centre. I mean you cannot waste your time on these things.	
<b>Peer or Self-assessment</b>	I would rather use this peer assessment or self-review for their developments, not for grades. I would give it as a formative assessment, like checklists or using the rubrics on their work or their peers work, but not necessarily giving them the grades. We have to train students to use rubrics to assess their work and others. We may use error codes, so we have to train the students if you want to involve them in this process.	I believe that they should be involved in the assessment process. I want them to get familiarized with assessment s. It is not something that they should fear.	We could use self-reflection and peer assessment during presentations; it does not have to be written. We could give a grade on peer-reviewing. Yes. I am presenting; they are grading me. It would be giving a purpose, and they would be motivated to move to work on it more.	I might give students rubrics, and then they can assess each other's based on that., and for the self-assessment, I could be based on observation by keeping a portfolio.	Peer assessment would be part of the syllabus. It should be integrated within all language skills. Peer assessment will happen when students listen to each other and discuss, reflect on the way they speak. They will learn, but again, not all teachers have this ability; they need training. Before we implement it, we have to have techniques. Here we do not need theoretical training at all. We need tangible techniques. Self-assessment is a further step, but it needs highly advanced students. I will not apply it with all students.	Honestly, because I used to apply this before in different institutions, I have a peer assessment rubric, through which students assess each other. I can provide them with that rubric; it is so detailed that it allows them to identify mistakes and correct them or have a discussion, a correction discussion for correcting these mistakes. So, it was very good and very successful. And in this case, with this formative based assessment program, it is the core of what we are doing. So, if your assessment is going to include all teachers, include students' voice, and link it to objectives, then it is task-based, including a variety of assessments. All of these discussed ideas are of ethical principles of assessment because this is the ethicality and the	Since I want to incorporate two different projects going at the same time, I can use these tiny bits of the tasks they do as peers or self-assessments. Let us say they have to prepare something like a piece of writing; they give it to the instructor. I give them feedback, and then they rewrite it for the second draft. For the second draft, I can use peer assessment and give their reports to each other. I can use self-assessment after my feedback as they write it again. But we have to wait for some time for this because I want them just to settle down, understand my feedback. I would keep it till the middle of the course until they know the system. Then I can ask them to self-assess and give their reports to me. So I will put everything in there. Maybe for peer assessment, We can create rubrics together in the class, where I can	I do believe that students have to work with each other because sometimes students understand from their friends more than they understand from the teachers. After all, they might have the same level of understanding. But as a teacher, I am the one that has to assess the student as at the end. I do not rely on the student to assess each other. Teachers are more qualified. Students are not up to that level to grade to assess each other. They might take it for fun. They might feel more confident that I am assessing my friend. See, I understand more

						<p>fairness. If you are giving them what they need and if you are judging the use of the language, this is ethicality.</p>	<p>train them first to use. It would be a chance to clarify why are you testing? What kind of objectives are you after? What do you want them to learn? I have to be careful that the rubric should be concise, clear, and understandable with the students.</p>	<p>than him. He is weak. I am good. Some students bully each other throughout this. So, I try to avoid that in an assessment. I use it as a builder, class builder or team builder, but not as an assessment.</p>
<p><b>Assessment Feedback</b></p>	<p>Students should submit their tasks to Turnitin. So, that they can see them all the time, they can create their folder, soft folder as well not only hard copies. They can also have soft copies; they can receive the feedback on online, they can see the feedback online, and they can work on it. So I think it would be more practical if we use error codes and we train students on comprehending them in addition to descriptive feedback online. Technology should be involved; we can use video conferences, and we can also leave audio feedback on</p>		<p>It is going to be ongoing.</p>	<p>So, I would always give continuous feedback in my classes, and it should be done weekly, individual meetings with students focusing on their weaknesses and how to improve that in the future. I would focus on the strength not only the weakness.</p>	<p>I need to involve that in skills, such as writing, reading, speaking, and everything. Students should be producing something. Teachers cannot be excluded from this. This is for teachers, but we need training, and I want to do it. It is important. It is the relationship between the teacher and the student. We need training in the different ways we can use to approach students, the reasons for using any of them. What to apply, when to apply and how to apply? What works? What does not work? So, that they will not feel bored, demotivated.</p>	<p>The feedback will be the major content of the assessment process. It is like you are assessing every stage, and you involve students in this feedback because you already give them the feedback that helps them to go on. Because assessment in this program is based on formative assessment, the key here will be feedback, addressing feedback, addressing strengths and weaknesses, working on the coming level, dividing work into tasks not into weeks and syllabus but on objectives. I finish this objective. I make sure students have acquired everything so that I would go to the next</p>	<p>I could keep a portfolio or a record. I will ask them to keep their work because I will not give them grades. I will give them feedback. So, the next task should be built on this feedback. Furthermore, I give another feedback for the next task, and the next one should be built on that feedback. So by the end of the semester, they will come up with two different portfolios So they will be working in different areas because not all students are the same. I am trying to develop their language skills. The function of English preparatory program should be to prepare learners to be university students. They should know how to study, responsibly</p>	<p>It is very good to give your feedback or to listen to the student's feedback about their assessments. It is very important to solve the test itself with the students giving them the answers, so they know which ones they got right or wrong. Moreover, when you get the results, sit with the students and discuss the feedback. For writing, for example, if I want to give feedback, I would not go for underlining the mistakes and use</p>

	<p>Turnitin. Some students might need some kind of extra feedback, like face to face feedback meetings, interviews, interaction with the instructor. That is why we have these writing centres, labs in colleges and universities. Those students should seek help from other resources on campus.</p>				<p>This one I need because it is this is part of my teaching and assessment tasks. You cannot isolate them.</p>	<p>step. So, it is about keeping records, monitoring, and developing learning. So, it is an assessment for learning not of learning.</p>	<p>organize, how to research, how to listen, how to comment, how to criticize. It has always been one of the most difficult tasks for teachers to work with foundation programs. Again, they are the frontline as they just received the raw material from the world, and they try to shape them into college students somehow and then pass them to their departments. So, it is not about grammar and vocabulary and reading and writing; they are the tiniest parts of their job.</p>	<p>error codes. I would add the correction of the word itself; some other people might have a point which I respect. They say it is the student's job, too. I believe that if students are too weak in English, and they do not even know how to use a dictionary. Well, I write the correction of the word itself. I would give it directly. I would also give commentary. When they read the comments; they know their mistakes to avoid them next time. I will conduct face to face meetings, definitely if the students' grades are really low. You need to sit together to find the problem. Nevertheless, I would advise them to go to other internet sources to help them on how to write, how to write</p>
--	--	--	--	--	---	--	--	--

								<p>in a better way, how to use sentence structure. For reading, I will give them sites, which provides them with some articles and stories to read to gain more knowledge to learn more vocabulary, learn more synonyms. Because sometimes they struggle with reading and they do not understand the questions because of hard words. For listening, I told you one of the tasks I would do is to check their listening while introducing any topic in class. I would give them commentary to be more fluent. Even for the listening, feedback could be on the topic because some students do not know how to summarise or paraphrase what they have heard. They go off-topic,</p>
--	--	--	--	--	--	--	--	--

								and they lose it all.
<b>Result Communication</b>	The results we get from the ongoing assessment will feed the system, in other words, we will use it to adjust teaching, develop learning and learners so that they can achieve their intended objective to be able to go to the following level. So, we will analyze their outcome, give feedback, follow up with learners, reassess the objectives. We would not let them sit for the exam unless we are quite sure that they acquired what is needed. Results from the proficiency tests will be analyzed to know the strength and the weakness of each learner; they need to go through more training based on their weak areas only until they reach the intended level, they do not need to repeat everything all over again. We can	We need to be trained and informed on how it would be best done.	There will be ongoing feedback given to them; we teach, assess, and give feedback. The whole process will be transparent.	It should be communicated in meetings	we need to be transparent about this it should be communicated to them in the form of feedback or a report for improving practice itself, for the whole process.	I would keep records compile and then come up, with a cumulative grade or a score at the end?	Feedback should be on writing, on everything. No grading. No scores. Verbal or written. You would base your teaching and assessment on your project. You would inter-relate them. So, from the very beginning, it is going to be a project you teach, you assess, and it is all coming under an umbrella of a certain project with a huge topic that may be related to the two main disciplines.	

	have individual meetings with them based on their exam results, and what questions they missed, what questions they should work on, what areas they are weak at. They have to know their pitfalls, their problems because we are expecting very few failing students after all this effort.							
<b>Validity &amp; Reliability Check</b>	If you were to adopt such a system, we have to do some regular checks even if they are classroom assessments, in the end, it should be there. We have a reliable tool to check if the students achieved those objectives. We have to have something standard in the end because our main goal is to make students reach that level of language. So we have to check if they reached that level in order to give them the pass. In the end, we have to assess their achievement.	Yes, with validity and reliability, I need an expert. I need to check. I need guidance. I need somebody to help me.	When I do this, I believe our assessments will be valid. I think you have to trust them. If you do not trust them, you would be in a situation where we are in right now. So, this is our place. It is not a sustainable environment. No one here will be happy.	If I do not have professional training, I would not make sure or confirm any of the above.	The consultant would manage this, but I need to have an idea. I need to understand. I cannot say that I am like paying 100 per cent attention to these things when I am creating because I do not have the time. They could give you the feedback from this. It is going to be a learning process on like a smaller scale one.	I guess since it is shared work, then all of us have a say, whether it is valid or not. And then again, if there is a committee to check its validity, then again that is something to be counted. They call this discussion validity because it is based on how you use and functionalize your learning outcome. In this case, that would be enough validity. If you are not using tests, you are not going to be using that statistical stuff. You will not need them.	So, you will go for formative projects, not for evaluation, Just to keep a record of how the student learnt and went on with the learning process. And when you believe that you have addressed all the objectives, you did your task, you gave the feedback, and you scaffold the students to reach whatsoever you had, then you will start creating the summative ones for the sake of grades and officially as a benchmark or as a gatekeeper. However, when you come to do this, you need another judgment a technical one, just to validate what you do because you believe you need a third eye. This	I think one of the problems nowadays is using the test more than once. Some teachers use it every year. I only modify or paraphrase the question. We need to be creative and come up with new tests. I will check this if I receive the appropriate training. I can create any assessment, but the problem is that I am not sure that my assessment was reliable. I want



							will take us through the following point. Teachers' involvement in assessment. You explain this already when you said that you want to involve specific teachers with the potentiality,	specialists to validate what I do.
<b>Fairness &amp; Ethical Procedures</b>	One of the aspects of fairness is to give people what they deserve or to allocate them where they are. You went for ethicality at the very beginning. You trained them. You satisfied your conscience by making sure that they acquired everything. You made them involved by assessing themselves and others; so actually, you maintained ethicality and fairness throughout. However, the question that crossed my mind. If the assessment is subjective and if it is based on learning and achieving objectives only without any numbering, scoring and rubrics, then we could be under	I would use an expert with fairness and ethical procedures to guide me. I would be happy about that. I will be confident if I have such guidance. I need to depend on a valid rigid standard tool whether standardize d tests or not; I am after any tool that would give fair results or fair assessment s and so on. I would use those other alternative assessment	We talked about ethicality, and the first step in ethicality and fairness is to be fair with the teacher and the students. Step one is a placement test. I do not believe that a final exam, one final exam should not allow the student to leave the program; we need alternative assessments. We should give them proficiency test every three months; if they can pass that proficiency exam, let them go to the undergrad. I feel that whatever I have put here now is fair and ethical to the student. I guess it would be quite satisfied with this; it may not be perfect; there is, of course, a room for great improvement, but it would be a start. It would be a good start.	It all goes back to the point of professional training and knowledge about assessment. So again, we are going back to this source of the problem.	I feel that in this part, in particular, I need more and more but this, in particular, have to be given to all teacher. I need to know how to implement it, how to make assessment fair. Ask any teacher. What is the meaning of being fair? You will have different answers. Mostly it will not be related to the context, the content of the assessment or how the assessment should be. It would be about their characters, about themselves. We should have continuous training about this because it changes from time to time. After all, many variables are involved in it.	We can as much as possible variate the levels of tasks and in the rubric, make sure that it serves this purpose. I guess it is good we will assure learning, give voice to students because we are going to involve peer on self-assessment in our assessment.	When we are talking about fairness and ethicality in the project, just setting the purpose is a fair practice. Helping the students and taking their hands and making them aware of what they are doing is ethical. Involving them is an ethical practice.  Listening to their voices is ethical, involving teachers is ethical, listening to teachers, ethical, giving them the right. Having this starting and ending with the professional technical assistant is ethical existence. So, ethicality is not just my test is fair. My test is ethical, calculate my validity, and calculate my reliability score. This is not only ethical, and you have used different assessment methods. This is ethical. You have addressed the different styles and strategies; this is ethicality. And as you mentioned, now if you are	

	<p>question; that is why I came up with the idea of the proficiency test at the end, we train them, we make them sit for the test.</p> <p>Students have come with such background where the main idea and focus is the score and the final grade. So now it is difficult to break this habit of focusing on grades and focus on evaluating; there is a pressure going on here. Teachers' opinion would be one big obstacle because based on my experience so far in this context, it is very difficult to reach objectivity. When it comes to teachers, when it comes to students, when it comes to real satisfaction, it is difficult to satisfy the students. It is difficult to satisfy the teacher. It is difficult to satisfy the administration because everybody is afraid. That is why it is so difficult</p>	<p>s as a teaching tool, but not as assessment tools. I want something concrete to make sure that teaching took place, that there was a record of it, that there is a document that proves it, that that there is something to depend on to move learners to the next level, a report, a record that says they know, it is like a driving license. There are some very difficult cultural issues to alter. We have a context where we</p>					<p>going to go for objectivity, right, you have this guy who is there for you at the very beginning.</p>	
--	---	---	--	--	--	--	--	--

	for teachers to be fair or to be objective, and they prefer to be lenient. I am not saying that they should be stricter and they should penalize the students for every mistake. The whole change is against this idea of penalizing students. I think maybe the only way to implement such a big change would be still having some kind of formal exams going together with the main ones. I think the placement and the exit exam should be better for the program. There must be some kind of paper documentation of the grades in order to defend yourself in front of students, in front of teachers, in front of the system.	cannot just ignore. We have students, parents, cultural issues, a native language that is used in the class, so it is different. Tests are important. I mean, they should be designed in the right way, but we cannot completely ignore them because of the context we are working.						
<b>Comparison</b>	Then if I had this dream system, so many good adjectives would come to my mind, I would say: Student-Centered, Learning-Oriented, and More	What we are doing is just, intuition, we are doing it by practice. By experience.	There is a lot to eliminate from my current context: there is too much going on; the number of deliverables, tests, quizzes. They tried to move away from the midterm and final	Since we started, I mentioned some sort of dissatisfaction about my current context. And then when it came to solutions, I gave some solutions. However, all of these		If I am going to describe my context, I will describe it as solid. It is rigid. It is not flexible, and it does not allow for freedom of manoeuvring and manipulating or	My dream practices will be in rainbow colours. It will be in rainbow colours, sparkly rainbow colours, I am seeing all the students happy and finding something for themselves in that	My current practices: not bad, and my dream. Unrealistic. It is because we always look for the ideal things, you know? I have

	<p>Democratic More Open-Minded. Very Satisfactory from my side, And Very Organized especially if I compare it to the chaos now, because we will start the whole thing with a very organized way in a very organized way.</p>	<p>Not very formal, not very scientific. Something that is done based on our experience as teachers, not very confident that this is how it has to be done. We have tried our best, but again, we are not provided with enough training. What is on the other hand, this one, I would be very confident of. I would be very knowledgeable of. I would do it as they say by the book, or I would learn it the right way. It is not. It is no longer trial and</p>	<p>exams, but they put the extra load, extra tests. You cannot move away from a test by giving another test. For the current one, I would say there is room for great improvement. Moreover, I would say that I would be a lot closer to a fair and ethical procedure, and it would benefit both the instructors and students. I will be a lot more satisfied.</p>	<p>solutions could not be implemented because I do not know how to implement them. These are all ideas. It can never be implemented without having professional workshops and technical way of doing it; even those high hopes are pointless without having the necessary tools, knowledge and experience. I can say that I have done my job fairly. I have this professional training and expertise and professional knowledge in practice. I would say I have done my job in a right, fair, ethical way. My dream scenario is not based on intuition.</p>		<p>moulding. It is just the mould that you cannot change. It is just one shape that you cannot change. Moreover, I would also say for the part of the dream assessment that it is a dream. I know it cannot be applied although I believe in it. Some parts of it were applied in different organizations, but not all of them. Because I suggested things like that before or tasks like that before, but they did not. Maybe because of the power of adopted assessment, maybe because of the motives of ideas, maybe because there is no awareness. Alternatively, if there is awareness, there is some kind of domination of certain textbooks. There is also political and economic power associated with language teaching and assessment, which is the power of domination of certain ideologies. Moreover, the books that come from certain parts of the world belong to</p>	<p>practice, but my current practice is basically grey because there is a certain path, there is a certain task, and we all have to abide by it.</p> <p>In my dream practice, we are free. We are happy because then we are creating ideas. We are creating thoughts. I mean, I am reaching their minds. I have the ability then to mould them. They are living the education experience; they are alive. They are having the experience of their lives there and me, too, and I am reliving living your life once and twice a year. In my current one, we know it is like a machine. I feel like a machine like I have to. There are certain deadlines that I have to keep to the deadlines. I have to give them that assignment, that assessment because I have to enter the grades, I keep entering the grades. You enter one set of grades three times, you know, and I have two hundred students. Imagine me entering grades. This is apart from excel sheets; I am wasting my time, I am wasting my eyes, I am</p>	<p>to mention that assessments have to be fair because I feel with the student. They sometimes study hard, but when it comes to the assessment, they feel down, and they get frustrated that they get low. We need to figure out what the problem is. That is a problem. The problem is following ethical procedures when we come up with any assessment tools.</p>
--	--	--	--	---	--	--	--	---

		error. This is how it has to be done. This is how I am going and doing it. I would feel very confident about the results because the results would reflect whatever is happening.				these ideologies and vice versa. So, there is no way to do it the way I want.	wasting my energy. By the time I go to class, I am already drained. So this is the current practice.	
--	--	---	--	--	--	---	--	--

## Appendix 17: Analysis of Data Collected about the Participants' Experience with the Research

### Activities

Criteria	Table 3.7 Participants Response Regarding their Assessment Literacy Development							
	Mirjana	Jasmine	Tok	Sherry	Rose	Rawan	Janset	Talen
<b>Experience with the Open-Ended Questionnaire</b>	No comment	No comment	No comment	It relates to knowing our current practices and if we were qualified enough to do such a role in terms of assessing students.	No comment	No comment	It was an excellent opportunity for reflection. It made me remember what I learned, how I was trained because, in this hassle of life, you just let loose tips.	No comment
<b>Experience with the Open-Ended Checklist</b>	It gave me some kind of awareness of what we should do. It also gave some confidence about what we were doing, although we did it without knowing that it is correct or right according to the book. I have realised that I do not know much about the theoretical part of assessments, whereas I do implement things in practice without knowing. I would appreciate some kind of support from the management in terms of assessment preparation, design, and evaluation process in general. Before I thought that assessment evaluation through statistics is very important, and we need it badly. Now I feel we do not need that much of details. I think the most important part is that we should change the whole system. We need to make a radical change and minimize the number of written assessments; let us say classical	I liked the fact that it did not only focus on theory, but it also focused on practices and steps and other important details, which I was not aware of. I felt it is holistic and thorough.	There were parts where I felt I need some polishing especially related to validity, reliability, statistics, so it was some sort of self-evaluation	I gained knowledge from the checklist; I also had the chance to provide my own opinion. Moreover, it is kind of self-evaluation and self-reflection on things. I realised that I was doing many things guided by my teaching experience. However, I realised that I do not have sufficient knowledge, and I do not have sufficient training for doing	I felt that the checklist added insight because as teachers, we work hard, and we do not take time to reflect on our knowledge and skills. This one made me stop and assess every part of my assessment knowledge and skills. It helped me understand which skills I need to sharpen and which ones I need to add to my current ones. I was performing many skills intuitively without knowing. I realised that there is a bit more that I need to acquire to carry on my role as an	It gave a very good idea of what I should keep in mind. If I get to evaluate my assessment experience, I know now what to think of and whether it is useful or not.  That was very rich.	This was the most detailed study I have ever experienced. It was good.	It is very good to use on any assessment.

	<p>written assessments. Instead, we should come up with some new techniques and some less stressful ways for students to assess their learning and give them constant feedback about their improvement and involve them in the process because language is a process. We can only check the reliability and validity of exams if we use them at the end. At that point, I guess we will not need that much of statistical data if we do not depend on one exam result. The checklist gave me so many ideas because it was very detailed, and it had made me think about every detail that we are going through here. Our discussions regarding the checklist were very provoking; so it was fruitful. For me it a good reflection practice on what we are doing as assessors and teachers, what we should be doing, what we are missing. It is very detailed.</p>			<p>assessments. I found it very exhausting, but I find it very valuable as well.</p>	<p>assessor. It allowed me to reflect on what I was doing currently. It helped me assess myself as an assessor. I now know very well what I need to focus on. It gave me a voice about what should and should not be involved as part of assessor's knowledge.</p>			
<p><b>Experience with the semi-structured Interview</b></p>	<p>When you do these things, you are not aware of what you are doing. Once we started discussing the assessment types, and how we evaluate students and what we should do, what we should focus on, I started to see my actions and that certain things should change. I had a chance to reflect on what we do here. It was also a good experience for me. I can think about serious proposals for things we need to change certain things afterwards.</p>	<p>It gave me the chance to reflect on my practices. I felt we have been neglecting certain things. We can do things in a very different way, but we are not equipped with these tools.</p>	<p>It gave me voice and chance to reflect and talk about things I rarely have the chance to talk about.</p>	<p>It was a very good chance for reflection because I think the kind of facts that I mentioned made me feel guilty. I felt I am not doing a good job. Moreover, if I have these necessary tools, I would be fairer. What I was doing was not based on knowledge or training. I think somehow your question made me feel I am not satisfied with what I am doing. I should</p>	<p>I felt this was necessary because it gave me time to speak up for everything that I felt related to my current practice. No one is asking about these certain things, but they should be considered in order to develop any practice in any educational environment. This allowed me to express all good things and difficulties that I am facing, which should be changed for the future.</p>	<p>It was very interesting because I talked about what happens and what I hope for. No one asked us what we hoped for. Everyone expected us to do what we are supposed to be doing to follow the path. When you are asked about what you want to do and what is wrong about what you are doing</p>	<p>it is relieving and disturbing because I know they should not be done that way. I thought that if I keep them to myself, it will be less embarrassing, but when I talked about these things, I felt less disturbed. It was relieving as if they were off my chest.</p>	<p>I feel content to talk about my ideas, views, and how I can change them. I think if I am in charge of making any assessment now, I will take into consideration all the things which we spoke about, like ethical and professional things</p>

				do better. You compensated me when you asked me about my dream assessment practices. This somehow compensated this feeling of guilt, so in the future, I can work on getting professional training.				because many things were missing from the assessments that are done in my workplace. So it widened my knowledge about assessments
<b>Experience with The Report</b>	You let me just come up with my ideal world. You gave me the microphone to assist myself. It made me feel I could implement things; I could change. I have something to share. The problem is with the context; I am not alone. I cannot make such radical changes on my own. I can only propose these changes in our meetings. Maybe we can do gradual changes. We can start using portfolios with projects to see how learners' writing improves. If it comes up with proof that they are improving, we can propose some changes in writing assessments. So this way we can start slowly on a small scale. In this way, we can propose some changes in the whole assessment process because, in our context, it is difficult to make big changes or propose big changes. They can easily be rejected. So instead it is better to go slowly.	Taking part in this made me feel we can change a lot with simple steps, and it could be much better and effective and efficient. Moreover, I could reflect more on what is happening in teaching. I think it would be fair for teachers and students.	This utopian part is my favourite stage among all other stages. This experience made me put everything into more perspective; it also made me happy about myself that I could come up with these creative ideas. They need to be implemented, just to be provided with the opportunity, I would do a very good job. I believe	This is exciting. I think higher management in any place should involve teachers. It is a shame they are not doing that. It is also a shame that they are not giving the teachers workshop on assessment, because again if they want to deliver safe practices for students, they have to fix this problem. The problem will be not be fixed unless it starts with teachers. So, I think it gave a little bit of space to talk about my plans if I am given a choice to do so. Well, it touched a nerve in me about something that I	It helped me put all the pieces of the puzzle together and then to think of practical solutions. I thought as if I were in a supervisor or a decision-maker position. I suggested solutions that depend on having guidance. We reached these conclusions with the different phases of the research.	I think this self-reflection, as much as it is also provided answers to my questions, it is also very important for instructors to get them to think about the solid models that they are trapped in, try to break it and create their satisfactory models.	It makes me remember why I became a teacher in the first place. Because I simply want to make a change in people's lives as simple as this. When I talked about it, I found that I came up with an amazing project on the spot; I did not expect that. I just in-cooperated all the energy that was charged throughout the research stages and poured it out in	It was a dream which I lived happily.



			there is always room for improvement	really should work on in the future.			this stage. I got carried away. You just opened me up to share my ideas, and I came up with an amazing project. I hope I would take this change take and incorporate it into something tangible.	
<b>Impact of such research experience on your assessment literacy</b>	It was a good reflection on what I have been doing for 15 years and what could be done in a better way. It helped me see that certain things are not necessary, whereas other things are necessary and they have to be there. I am happy to be involved in that because I had a very good experience thinking about what we do here and being critical about the things that we are doing in this context and how we can improve things in terms of assessments. So it might give me the chance to propose some changes in the future, on how to improve the assessment mechanism in this context. I have never been in a context where I was trained as an assessor. Now, when I talked about these things, it is interesting to me. I would want to know more. I felt the need for some kind of training about it. I want to know more about how people are doing. Such kind of contextualized research, makes teachers discuss problems, suggest solutions, be open to feedback and learning from colleagues. It gives voice to teachers. This is how we learn and improve. It is not necessarily that we could find a solution. We	It is an eye-opener kind of experience for me. I realised there is not enough focus given to a tool that decides students' future. Those numbers, tests, exams are very important, but doing it right is something that is neglected because we are focusing on certain formalities. I felt I want to know more. There are many things that I am not putting into consideration. I am doing things, but I am not 100 per cent sure I am doing it the right way. I started to have many questions in mind that require answers. A lot is going on in the field, and we are just stuck with formalities and details. I	I get to re-evaluate myself going through these stages, but it is also upsetting because I know that none of these things can be applied to the current context I am in. Yes, it gave me a push, but unfortunately, the context I am in does not provide me with that kind of opportunity. This kind of research	It gave me at least some background on assessment. The research is very informative because it provides some assessment knowledge that I was not aware of before, such as scoring, scales, steps of designing, different types of assessment. After the research, I knew it is very important for teachers to have this knowledge, whether theoretical or practical one before performing their assessment roles. We need to know how to assess the students in a	It helped me acquire some theoretical knowledge based on our discussion during the different phases of the research, especially the checklist. I did not know some of the terms. I used to apply some of them blindly. I realised that I should know about these things. It helped me reflect on my assessments and understand the whole situation around me. Now I understand that I was just judging students all the time. I realised that I need to have a counsellor to formalize our assessments.	I appreciated participating in this research because I learned many things or at least got to know that there are things that I think of and they exist. It gave me a better vision of what should be done or what might be done later if there is some kind of change. I will try to be more careful in the future when I am devising questions. I hope that I would be	I started to believe even more that we do not need this mode of assessment; it is too much for us as teachers and for students as well. I realised that whatever we were doing in terms of assessment was haphazard. Next time I am preparing for the exams, I will be more cautious. This participation shed light on my path. I now look at	It opened my eyes to different things. I felt that this research is detailed, and there were a lot of methods and criteria of testing and assessing, which I was not aware of. However, now I think I should consider them whenever I do any assessments. It was detailed and comprehensi

	do not have teachers' room anymore, but in those rooms, whole informal teacher development used to take place. Just complaining about things and trying to assist one another with solutions solved many problems. Sometimes you feel relaxed if it is everybody's problem, not just you, so instead of blaming oneself, you start supporting each other. Teachers usually blame themselves and their lack of experience, so we try to find solutions to our problems because we are on the same boat. Solutions come from inside; it can come from outside but to a certain degree. I can learn about the theory and how to prepare an exam, but I will prepare it in my context with this cultural background and for these students.	feel I need training, and I need to know more as an assessor. I have to have the tools to review what I have done and what I should do. There is some kind of a gap, and it has to be fixed somehow. Trial and error will not work anymore. Participation in the research gave me such awareness. I feel we need more standardization, more guidelines. I feel I am missing out certain things.	might be an eye-opener, and it might help me to reflect, criticize, create, have high hopes, but it stops at this current stage, it will not contribute to teacher development and practices; it will be just ideas trapped in a box.	fair, ethical way. So the research was comprehensive and detailed. When I go to class or approach students with assessments next time, I have to be very careful. Even when I give students in-class activities, which are not graded, I need to be careful. I will start to question the purpose of my practices. I might ask for higher management to give us training, professional training. So, it was a scream for fair educational practices.		given the freedom to be more critical.	assessments from a different perspective.	ve. I now want to revisit all my previous assessment practices and evaluate them based on our discussion and what we have done through the research stages.
<b>Future Development and Enhancement of Assessment Literacy</b>	I definitely would have appreciated the opportunity of having a practical knowledge on where to start, how to start, why we are doing these things, how we can improve it. I need some kind of support from the literature that would support my ideas or my suggestions. If I create a proposal, I need to refer to certain studies, and I can prove my points. I can argue more strongly when it comes to the proposal stage. I need to be confident about concepts. Technically, we need some training and on how to grade and how to be objective. Students should also be somehow be trained in the classroom to read through rubrics, to receive feedback, to work on the feedback, to change certain things. To	I might list now after this research: creating exams and assessment tools as areas of improvement that I need as part of my professional development requirement. I will try to communicate this with my department that we need some training for us as assessors or whoever is going to assess. I believe that this should include all teachers, but	Only one action will help assessors. They need training on how to assess, what to do, what steps to follow. We should take the research outcome and change it into something;	Now, after this research, I think if I am being asked to create assessments, I will refuse because again, it is very complex, and it is not an easy job to do. I will start working on myself. I will try to gain knowledge. I would never get involved unless I have this kind of training. It	I need to find practical workshops or guidance that could develop my skills. I need to suggest to my supervisors having a counsellor that could teach us different skills or validate our current ones. I started to appreciate peers feedback, especially experienced ones. We need to support one another. Moreover, it	I need assistance. I need freedom. I am not for those designs and those tests. I am not teaching chemistry or physics in which they have to memorize certain formulas and certain methods that should be applied. I am teaching a language, and the	I am not ashamed to say it. We need this third eye; we need assistance. I need a follow up on the findings of that research. For the parts where I said, I am not confident. I need assistance. I	I would try to make my assessments more valid. I will try to check more sources. I am now more interested in areas such as fairness, justice, which I want to explore more. I need workshops. I

	<p>improve their writing, they should understand the process. Otherwise, it would be unclear to them, and they will not accept it. After being trained; everything should start from the context. It could not come from outside. Everything should be shifted from here. We cannot even find a book to satisfy our writing needs. When it comes to exams, they have to be contextualized. Everything should be internal.</p>	<p>assessors need help. So, this is something that I would ask for.</p>	<p>we should try to do something effective technical practice.</p>	<p>should not be based on intuition or just experience. It should be based on knowledge and practice.</p>	<p>changed my perspective of assessments. I got the chance to think that it is something far more important; in other words, I think of it as a tool of learning more than of measuring and evaluating. It should be dynamic and changeable. It should be different from one context to the other, or from one level to the other. I started to consider the ethicality of assessment in more in-depth. I did not use to look at the impact it had on students. Now I will start applying different techniques; I will try to get feedback from my students about assessment.</p>	<p>language is flexible. I do not believe in standardized tests. Teaching and assessment is one entity. I teach, and I assess. I want to survive without those complications of psychometric assessments, at least in my context.</p>	<p>do not want all of these hours that we spent in the research to go in vain. You just opened the door, but what is next?</p>	<p>need to communicate more with a community of professionals specialized in assessments</p>
--	---	---	--	---	---	---	--	--